NASA-CR-200976

17

Department of
ELECTRICAL ENGINEERING

UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIANA

SEMI-ANNUAL STATUS REPORT

Error Control Techniques for Satellite and Space Communications
NASA Grant Number NAG5-557

Principal Investigator:
Daniel J. Costello, Jr.

May 1996

Summary of Progress

In this report, we present the results of our recent work on turbo coding in two formats. Appendix A includes the overheads of a talk that has been given at four different locations over the last eight months: the Allerton Conference in Monticello, Illinois (October, 1995); Motorola Communications in Schaumburg, Illinois (February, 1996); NASA's Turbo Coding Conference at Southern University (February, 1996); and the IEEE Communication Theory Workshop in Destin, Florida (April, 1996). This presentation has received much favorable comment from the research community and has resulted in the full-length paper included as Appendix B. This paper has been accepted for publication in the special issue on coding and complexity of the IEEE Transactions on Information Theory to appear in November, 1996. Major contributions to this paper were made by Dr. Lance C. Perez, a postdoctoral research associate supported by the grant and by three students from the Swiss Federal Institute of Technology: Jan Seghers, Guido Meyerhaus, and Dieter Arnold. These students have worked with Drs. Costello and Perez on turbo coding research over most of the last year.

Appendix A

The Turbo-Coding Talk

A Distance Spectrum Interpretation of TURBO Codes[1]

Daniel J. Costello, Jr.
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Lance C. Perez
Division of Engineering
University of Texas at San Antonio
San Antonio, TX 78249

presented at the
IEEE Communication Theory Workshop
Destin, Florida

April 15, 1996

[1]This work was supported by NASA grant NAG5-557 and NSF grant NCR95-22939.

Introduction

- In 1993, a new coding scheme, called TURBO codes, was proposed for achieving near capacity performance in the power limited region of the *additive white Gaussian noise* (AWGN) channel.

- TURBO codes use a parallel concatenation of rate 1/2 convolutional encoders combined with iterative *maximum a posteriori probability* (MAP) decoding to achieve a bit error rate (BER) of 10^{-5} at a signal-to-noise-ratio (SNR) of only 0.7dB.

- The channel capacity for a rate 1/2 code with *binary phase-shift-keyed* (BPSK) modulation on the AWGN channel is 0dB, and thus the TURBO coding scheme comes within 0.7dB of capacity at a BER of 10^{-5}.

- In this paper, we present some results on the relationship between the structure of the TURBO encoder and the resulting distance spectrum of the code. These results provide an explanation for the excellent performance of this coding scheme.

TURBO Code Performance

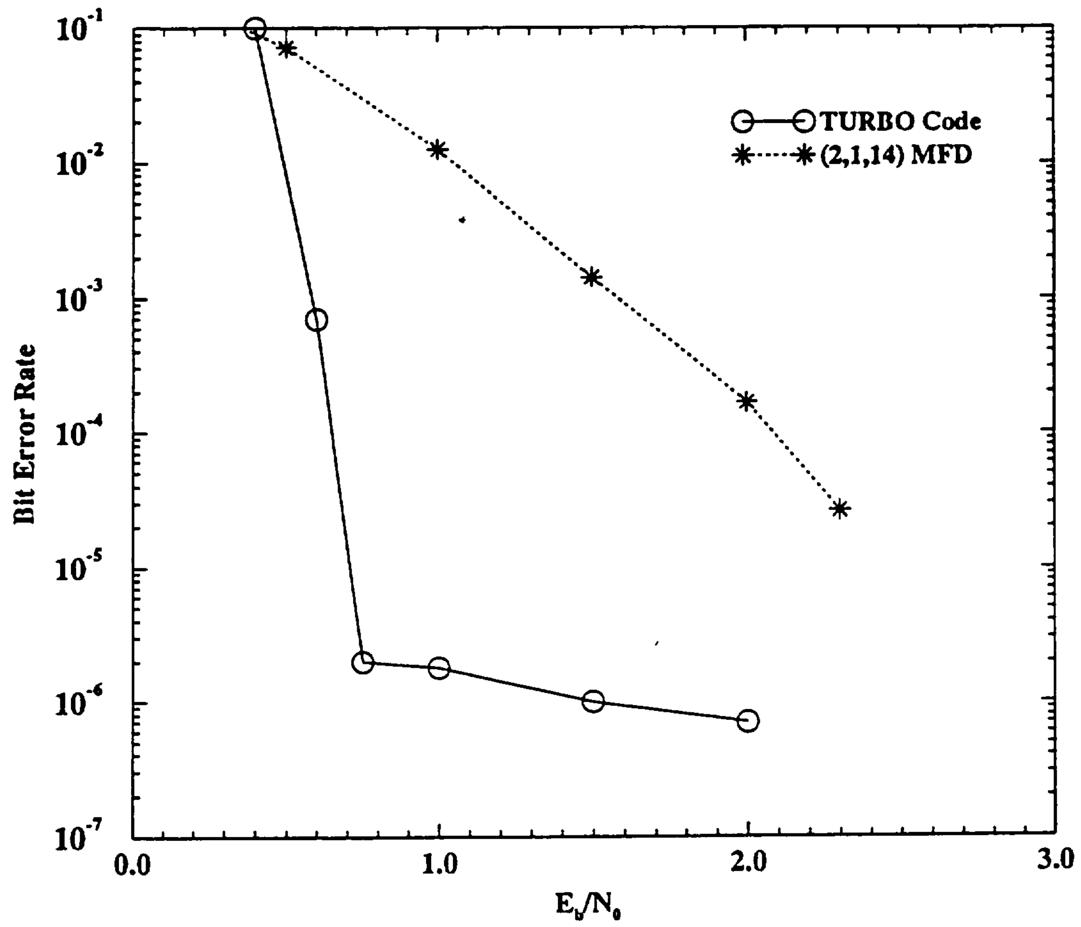
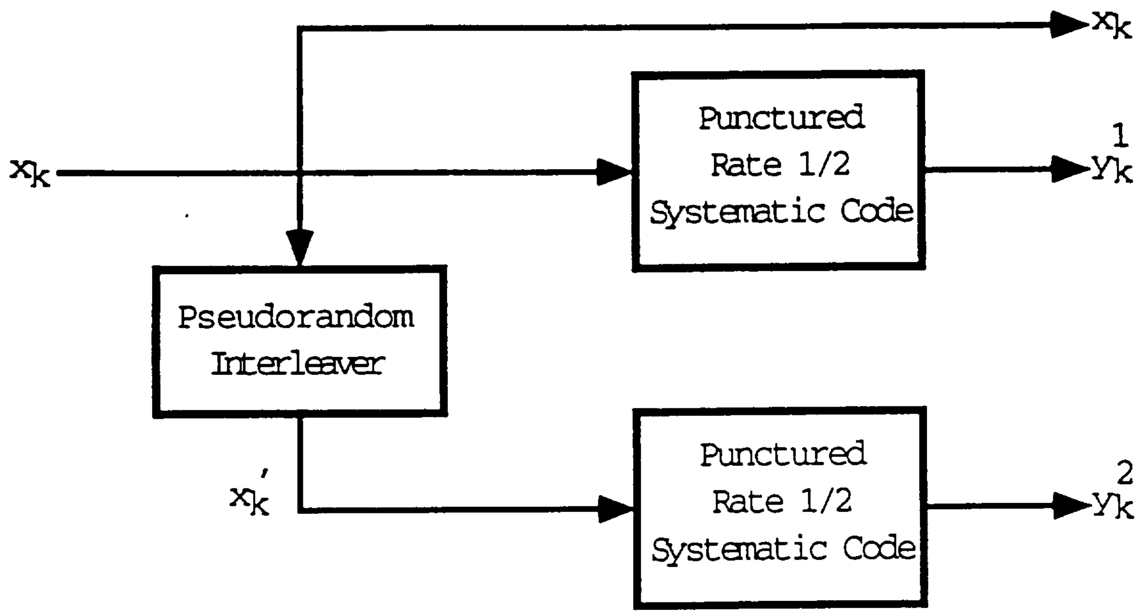


Figure 1: Performance of the TURBO Code.

- The performance curve of the TURBO code has two remarkable features:
  1. For low SNR's the curve is very steep, thus enabling near capacity performance at a BER of 10^{-5}.
  2. For moderate and high SNR's, the curve flattens out, resulting in what has been called the "error floor".
- The performance of the TURBO code is distinctly different from the performance of conventional convolutional codes such as the maximal free distance (MFD) $(2, 1, 14)$ code with Viterbi decoding.

TURBO Encoder



x_k
x_k
Punctured Rate 1/2 Systematic Code
y_k^1
Pseudorandom Interleaver
x_k'
Punctured Rate 1/2 Systematic Code
y_k^2

Figure 2: Block diagram of the TURBO encoder.

- The TURBO code uses two identical rate $R = 1/2$, constraint length $\nu = 4$, convolutional encoders in systematic feedback form in a "parallel concatenation" configuration.

- Each encoder is punctured to rate $R = 2/3$. The systematic form of the encoders results in each information bit being sent only once. Thus, the overall rate of the TURBO encoder is

$$R_{TURBO} = \frac{2}{2+1+1} = \frac{1}{2}.$$

- The pseudorandom interleaver insures, with high probability, that the codeword generated by the first encoder in response to the input sequence $\{x_k\}$ is different than the codeword generated by the second encoder in response to the interleaved input sequence $\{x_k'\}$.

- For example, the encoder used in the original TURBO code has the following generator matrix

$$G(D) = \left[\, 1 \;\; \frac{1+D^4}{1+D+D^2+D^3+D^4} \,\right].$$

- If the input sequence is $x(D) = 1 + D + D^2 + D^3 + D^4$, then the output of the first encoder is $[x(D), y_1(D)]$, where $y_1(D) = 1 + D^4$ is the parity output (ignoring puncturing).

- Suppose the interleaver maps the input sequence $x(D)$ to the sequence $x'(D) = 1 + D^2 + D^3 + D^5 + D^6$, which becomes the input to the second encoder. (Note that the weight of the input sequence is not changed by the interleaver.) The output of the second encoder is then $[x'(D), y_2(D)]$, where $y_2(D)$, the parity output, now has infinite weight, since $x'(D)$ and the denominator polynomial $1 + D + D^2 + D^3 + D^4$ are relatively prime.

- In this manner, the pseudorandom interleaver makes it unlikely that both encoders will generate low weight parity sequences in response to a particular input sequence.

- It is tempting at this point to conclude that the excellent performance of the TURBO code is due to a large free distance. However, this turns out not to be the case.

TURBO Decoder

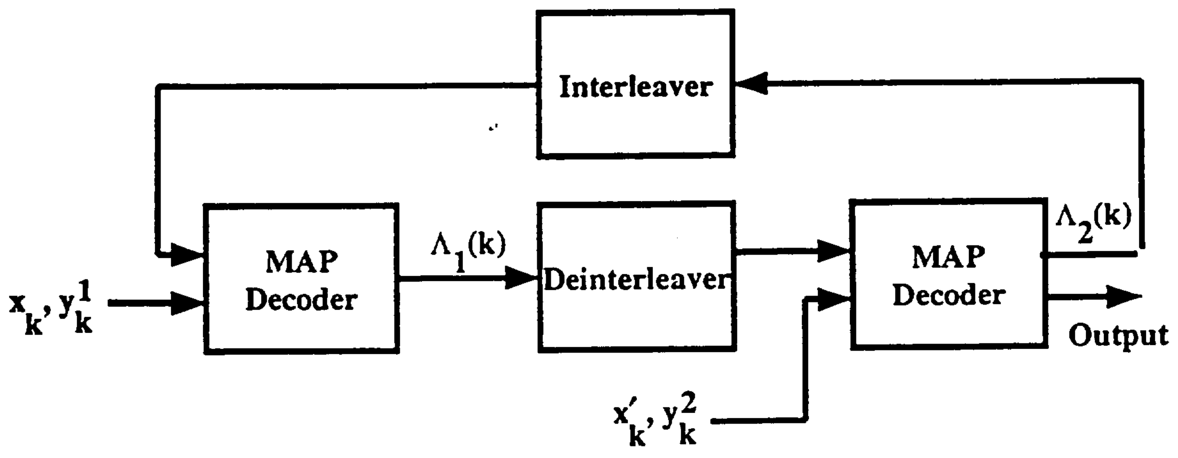


Figure 3: Block diagram of the TURBO Decoder.

- The "parallel concatenation" of encoders is iteratively decoded by two identical MAP decoders.

- In order for the iterative decoding technique to be effective, soft information must be passed from one decoder to the next. This is done using a portion of the log-likelihood ratio, $\Lambda_i(k)$, calculated by the MAP decoding algorithm.

- Each pass through the two decoders counts as one iteration.

- A total of 18 iterations are required to achieve a BER of 10^{-5} at an SNR of 0.7dB.

Convolutional Codes

- In order to understand the performance of TURBO codes, it is instructive to first consider the performance of convolutional codes with finite length input sequences and maximum likelihood (ML) decoding.

- With finite length input sequences of length N, a $(2, 1, \nu)$ convolutional code may be viewed as a block code with 2^N codewords of length $2(\nu + N)$.

- The information bit error probability of the code is upper bounded by

$$P_b \leq \sum_{i=1}^{2^N-1} \frac{w_i}{N} Q\left(\sqrt{d_i \frac{RE_b}{2N_0}}\right),$$

where w_i and d_i are the information weight and total Hamming weight, respectively, of the i^{th} codeword.

- Collecting codewords of the same total Hamming weight and defining the average information weight per codeword as

$$\tilde{w}_d = \frac{w_d}{N_d},$$

where w_d is the total information weight of all codewords of weight d and N_d is the number (multiplicity) of codewords of weight d, yields

$$P_b \leq \sum_{d=d_{free}}^{\infty} \frac{N_d \tilde{w}_d}{N} Q\left(\sqrt{d \frac{RE_b}{2N_0}}\right).$$

Convolutional Codes (cont.)

- If a time-invariant convolutional code has N_d^0 codewords of weight d and length l caused by information sequences $x(D)$ whose first one occurs at time 0, then it also has N_d^0 codewords of weight d and length l caused by information sequences $Dx(D)$, N_d^0 codewords of weight d and length l caused by information sequences $D^2x(D)$, and so on.

- Thus, as the length of the information sequences increases we have
$$\lim_{N\to\infty} N_d \to N_d^0 N, \text{ and } \frac{N_d}{N} = N_d^0$$
and the bound on the BER of a convolutional code with ML decoding becomes
$$P_b \leq \sum_{d=d_{free}}^{\infty} \frac{N_d \tilde{w}_d}{N} Q\left(\sqrt{d\frac{RE_b}{2N_0}}\right) \to \sum_{d=d_{free}}^{\infty} N_d^0 \tilde{w}_d Q\left(\sqrt{d\frac{RE_b}{2N_0}}\right).$$
This bound is dominated by the first term for moderate and high SNR's.

- For this reason, most efforts to find good convolutional codes have focused on finding convolutional codes that maximize the free distance d_{free} and minimize the multiplicity N_{free}^0 for a given rate and constraint length.

The MFD (2,1,14) Code

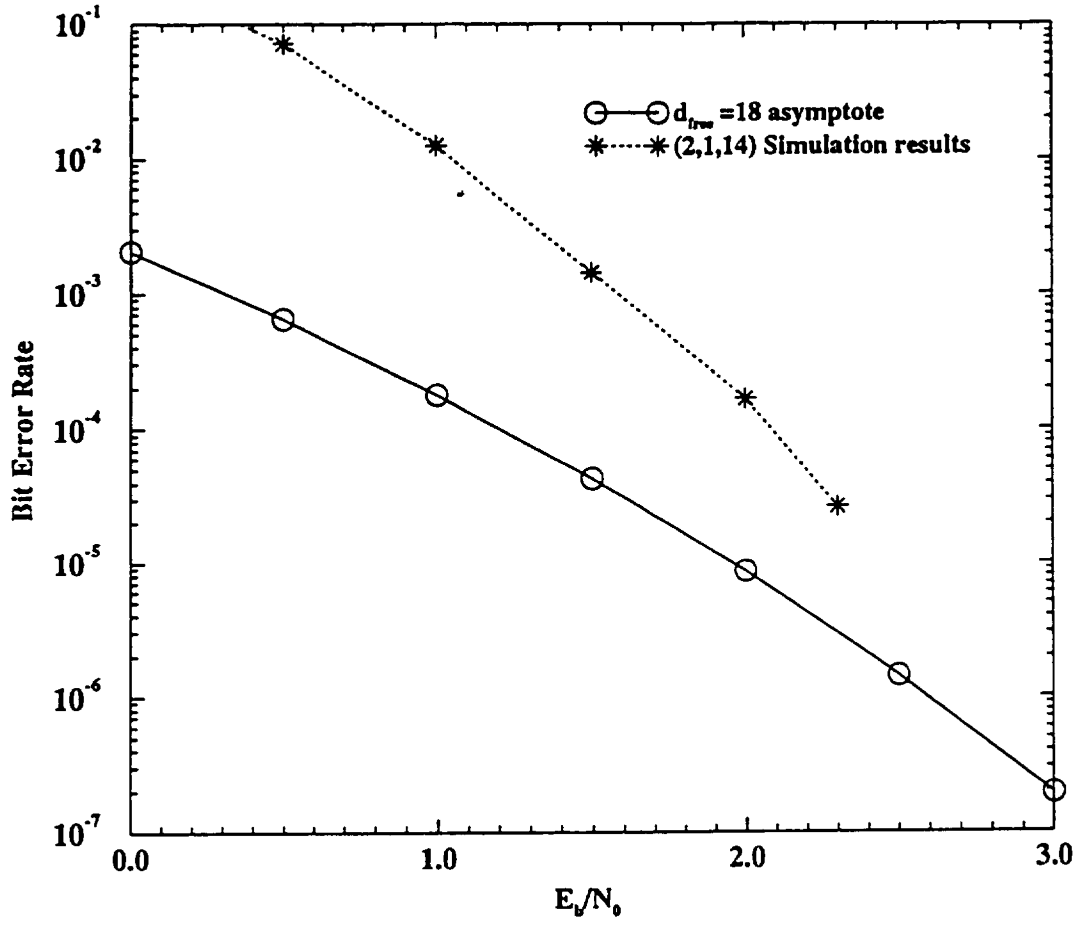


Figure 4: Asymptotic performance of the $(2,1,14)$ code.

- The first term in the bound is given by

$$N^0_{free}\tilde{w}_{free}\, Q\left(\sqrt{d_{free}\frac{RE_b}{2N_0}}\right) = w^0_{free}\, Q\left(\sqrt{d_{free}\frac{RE_b}{2N_0}}\right),$$

where $\tilde{w}_{free} = w^0_{free}/N^0_{free}$ follows from time invariance. (For the MFD (2, 1, 14) code, $d_{free} = 18, N^0_{free} = 33$, and $w^0_{free} = 187$.) This term is referred to as the *free distance asymptote* and accurately predicts the performance of the code for high SNR's.

- For low SNR's, however, the performance is much worse than the free distance asymptote.

The (2,1,14) Code (cont.)

- The performance of the MFD $(2,1,14)$ code for high SNR's is limited by its free distance asymptote.
- But for low and moderate SNR's there is a significant gap between the asymptote and the actual performance of the code. This gap can be explained by examining the distance spectrum of the $(2,1,14)$ code.
- The distance spectrum of this code is

d	N_d^0	w_d^0
18	33	187
20	136	1034
22	835	7857
24	4787	53994
26	27941	361762
28	162513	2374453
30	945570	15452996
32	5523544	99659236

- Recall that the total multiplicity $N_d \to N_d^0 N$ for large N!

The (2,1,14) Code (cont.)

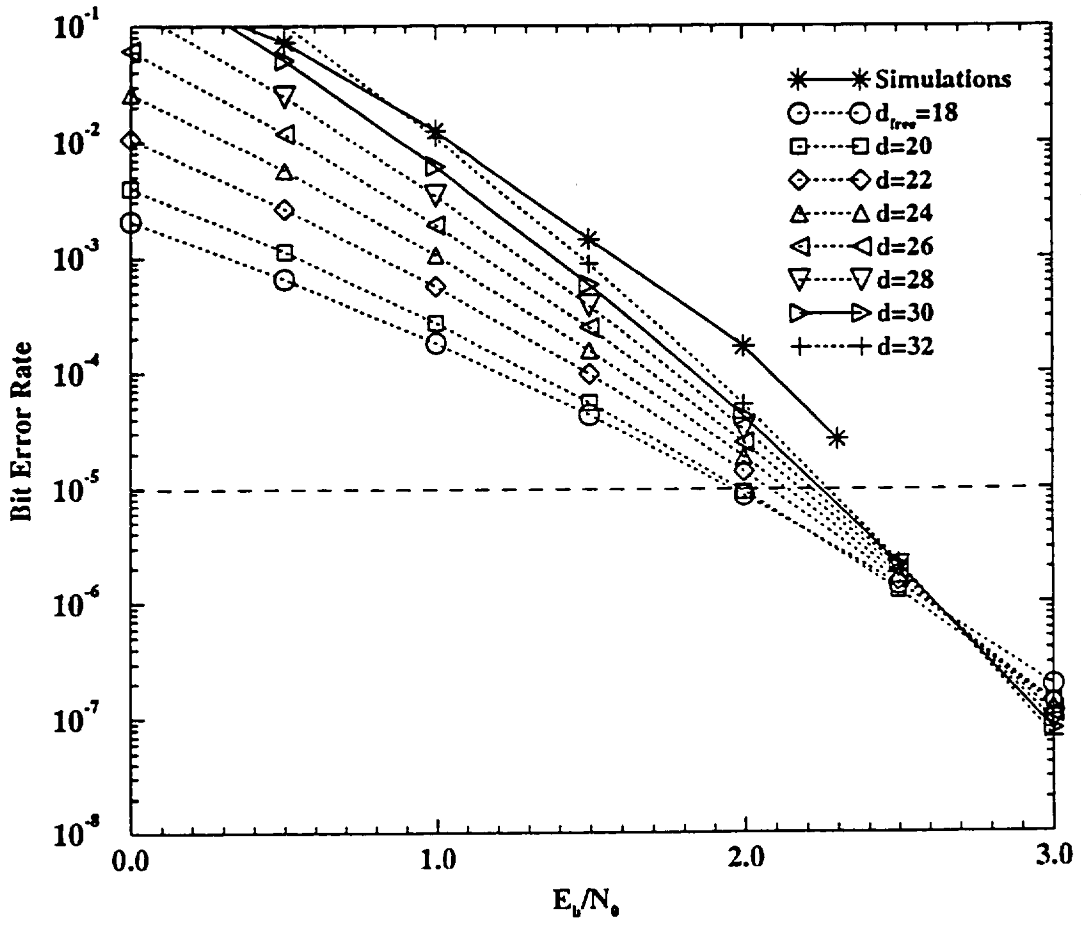


Figure 5: Decomposed performance of the MFD (2, 1, 14) code.

- By plotting the contribution to the BER of each spectral line, it is easily seen that for SNR's less than $E_b/N_0 = 2.7$ dB the performance is not dominated by the free distance asymptote.

- Instead, the higher distance paths dominate the performance for these SNR's due to their very large multiplicities.

- Thus, the relatively large difference between the real coding gain and the asymptotic coding gain of the (2, 1, 14) code is due to its very dense distance spectrum.

TURBO Codes

- The performance of the TURBO code with ML decoding is also upper bounded by

$$P_b \leq \sum_{i=1}^{2^N-1} \frac{w_i}{N} Q\left(\sqrt{d_i \frac{RE_b}{2N_0}}\right),$$

where w_i and d_i are the information weight and total Hamming weight, respectively, of the i^{th} codeword.

- However, in the TURBO encoder, the pseudorandom interleaver maps the sequence $x(D)$ to $x'(D)$ and the sequence $Dx(D)$ to a sequence $x''(D)$ that is different from $Dx'(D)$ with very high probability.

- Thus, the input sequences $x(D)$ and $Dx(D)$ produce different codewords with different Hamming weights.

- As the interleaver size N increases, the total multiplicity of free distance codewords N_{free} (not just those starting at time 0) approaches a constant that is much less than N. (This is not true for rectangular interleavers!)

- Thus, the free distance asymptote for a TURBO code with a pseudorandom interleaver is

$$\frac{N_{free}\tilde{w}_{free}}{N} Q\left(\sqrt{d_{free} \frac{RE_b}{2N_0}}\right),$$

where

$$\lim_{N\to\infty} N_{free} << N.$$

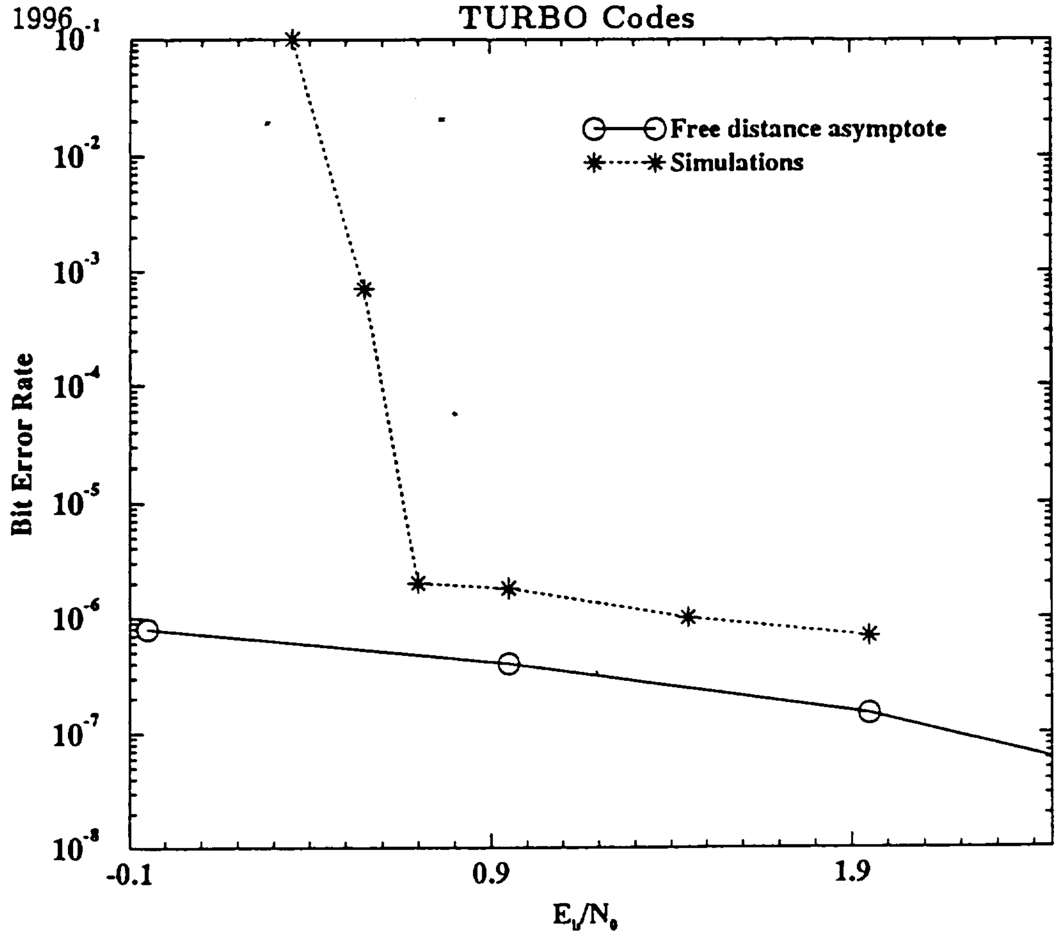


Figure 6: Asymptotic performance of the TURBO code.

- The original TURBO code with an $N = 65536$ pseudorandom interleaver has $d_{free} = 6$, $N_{free} = 3$, and $\tilde{w}_{free} = 2$.
- The free distance asymptote for the TURBO code is thus

$$\left(\frac{6}{65536}\right) Q\left(\sqrt{6\frac{0.25E_b}{N_0}}\right).$$

Note that the exponent is much smaller than for the (2,1,14) code, but the coefficient is also greatly reduced, resulting in a much flatter curve.

- Comparing the simulation results of the TURBO code with the free distance asymptote, it it clear that the "error floor" is simply the result of the code approaching its free distance asymptote.

Spectral Thinning

- In view of the analysis of the $(2, 1, 14)$ code, it is reasonable to suggest that the excellent performance of the TURBO code at low and moderate SNR's is due to a relatively "thin" distance spectrum.

- That is, the TURBO code is able to follow its free distance asymptote at lower SNR's because the multiplicities of higher weight codewords are small enough that the free distance asymptote remains the dominant term in the bound.

- The distance spectrum of TURBO codes is the result of a process called "spectral thinning" in which the interleaver effectively moves many lower weight codewords to a higher weight.

- This theory is supported by simulation results and actual calculations of the distance spectrum of several TURBO codes.

Spectral Thinning

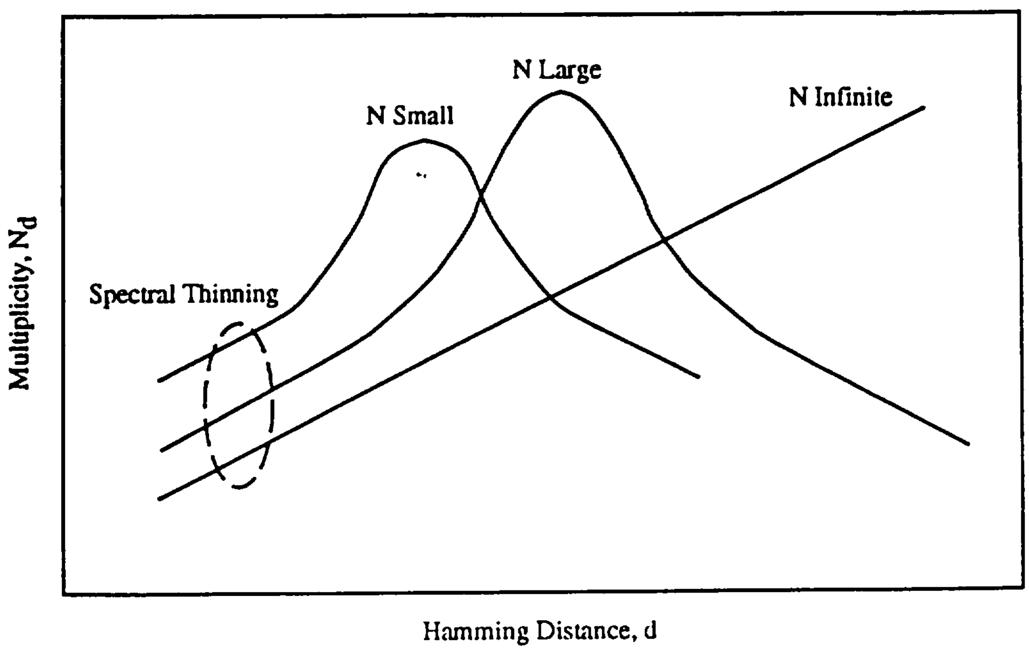


Figure 7: Hypothetical distance spectra of TURBO codes.

- This figure conceptually depicts the process of spectral thinning for three different size pseudorandom interleavers.

- As the size of the interleaver increases, more low weight codewords are moved to higher weights and the distance spectrum approaches a binomial distribution with small variance and mean close to N.

TURBO Codes (cont.)

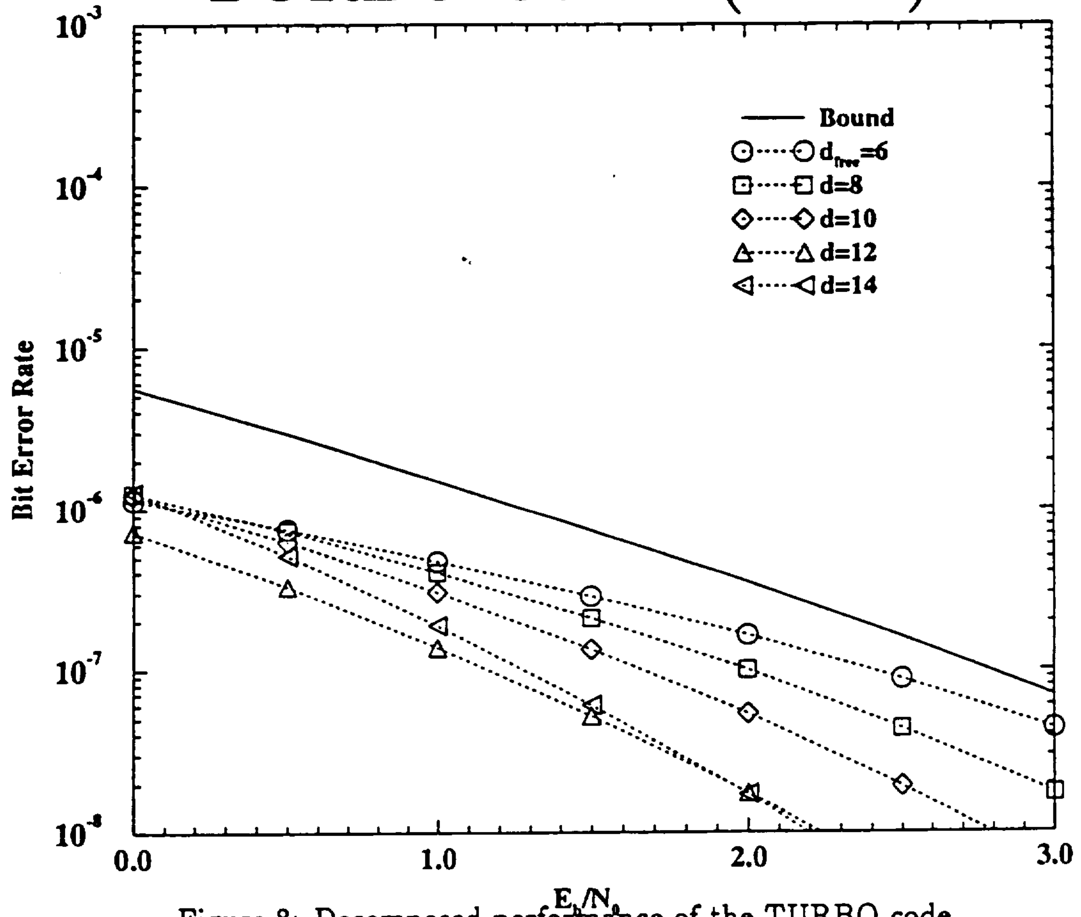


Figure 8: Decomposed performance of the TURBO code.

- With an interleaver of length $N = 100000$, the TURBO code has the following distance spectrum:

d	N_d	w_d
6	8	16
8	22	54
10	41	157
12	150	323
14	721	1462

Recall that these are the total multiplicities N_d, not just the values of N_d^0!

- In this case, we see that the free distance asymptote remains the dominant performance parameter even for low SNR's.

Conclusions

- The excellent performance of TURBO codes may be explained in terms of the distance spectrum of the code.

- The "error floor" observed in simulations of TURBO codes is a manifestation of the free distance asymptote. Since TURBO codes have relatively low free distances, the free distance asymptote dominates performance at moderate and high SNR's.

- The "error floor" can be lowered by increasing the size of the interleaver (without changing the free distance) or by increasing d_{free}.

- The exceptional performance of TURBO codes at low SNR's is due to "spectral thinning" and the resultant ability of the code to follow its free distance asymptote almost to channel capacity.

- The combination of long block lengths, N, and low multiplicity, N_{free}, results in a very small *effective multiplicity*

  $$N_{eff} = \frac{N_{free}}{N} << 1$$

  compared to convolutional codes, where $N_{eff} = N^0_{free} \geq 1$.

- The complete distance spectrum has a random-like distribution, thereby approximating a long, random block code that can be decoded with reasonable complexity. This is consistent with Shannon's noisy channel coding theorem and explains the near capacity performance at low SNR's.

Appendix B

The Turbo-Coding Paper

A Distance Spectrum Interpretation of Turbo Codes.*

Lance C. Perez
Division of Engineering
The University of Texas at San Antonio

Jan Seghers
Institute for Signal and Information Processing
Swiss Federal Institute of Technology
Zurich, SWITZERLAND

Daniel J. Costello, Jr.
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46616

IEEE Transactions on Information Theory
Special Issue on Coding and Complexity

Submitted: January, 1996
Revised: April 30, 1996

Abstract

The performance of Turbo codes is addressed by examining the code's distance spectrum. The 'error floor' that occurs at moderate signal-to-noise ratios is shown to be a consequence of the relatively low free distance of the code. It is also shown that the 'error floor' can be lowered by increasing the size of the interleaver without changing the free distance of the code. Alternatively, the free distance of the code may be increased by using primitive feedback polynomials. The excellent performance of Turbo codes at low signal-to-noise ratios is explained in terms of the distance spectrum. The interleaver in the Turbo encoder is shown to reduce the number of low weight codewords through a process called 'spectral thinning'. This thinned distance spectrum results in the free distance asymptote being the dominant performance parameter for low and moderate signal-to-noise ratios.

*This work was supported by NASA grant NAG 5-557, NSF grant NCR95-22939, and the Lockheed-Martin Corporation.

1 Introduction

The discovery of Turbo codes and the near capacity performance reported in [1] has stimulated a flurry of research efforts to fully understand this new coding scheme [2]-[44]. Initially greeted with some skepticism, the original results were independently reproduced by several researchers [6],[7], [11]–[12], [13], [14], and [31]. Subsequently, recent research on Turbo codes has focused on understanding the reasons for their outstanding performance [8]–[10], [16], [23]–[25], [29].

At this point, there are two fundamental questions regarding Turbo codes. First, does the iterative decoding scheme presented in [1] always converge to the optimum solution? Second, assuming optimum or near optimum decoding, why do the Turbo codes perform so well? In this paper, we attempt to address the second issue by examining the distance spectrum of Turbo codes. In doing so, we will draw on the work of several research groups involved with Turbo codes [6]–[10], [13]–[15], [19]–[25]. Due to the intense interest in this subject, many results involving Turbo codes have been developed independently by others and the reader is encouraged to peruse the references for an alternative point of view.

The simulated performance of a rate 1/2 Turbo code with the same parameters as in [1] is shown in Figure 1 along with simulation results for a rate $R = 1/2$, memory $\nu = 14$, convolutional code. The comparison of these simulation results raises two issues regarding the performance of Turbo codes. First, what is it that allows Turbo codes to achieve a bit error rate (BER) of 10^{-5} at a signal-to-noise ratio (SNR) of $E_b/N_0 = 0.7$ db which is only 0.7dB from the Shannon limit? Second, what causes the "error floor", that is, the flattening of the performance curve, for moderate to high SNR's? Here, we endeavor to explain the performance of Turbo codes, and thus address these two issues, in terms of the code's distance spectrum. We do not attempt to address the many interesting questions concerning the iterative decoding method (see, e.g., [27], [43], and [44]), i.e., we assume that an optimum or near optimum decoder is available.

In order to explain their performance in terms of the free distance and the distance spectrum, we will examine the codeword structure of Turbo codes in detail. Here, the free distance is defined to be the minimum Hamming weight of all possible codewords and the error coefficient is the total number, or multiplicity, of free distance codewords. The goal is to use specific examples to elucidate the key structural properties that result in the near capacity performance of Turbo codes at BER's around 10^{-5}. As will be seen, this effort leads to an interpretation that applies to Turbo codes and also lends insight into designing codes in general. Throughout the paper, Turbo codes are compared to a maximum free distance, rate $R = 1/2$, memory $\nu = 14$, i.e., a $(2, 1, 14)$, convolutional code to emphasize the differences in performance and structure. Techniques for analyzing the performance of Turbo codes using transfer functions and the like may be found in [6], [10], and [23].

The paper begins with a detailed examination of the structure of codewords in a Turbo code in Section II. This leads to the calculation of the free distance of a particular Turbo code and an explanation for the error floor in its performance curve. In Section III, the distance spectrum of Turbo codes is considered and a theory called spectral thinning is introduced and used to explain the performance of Turbo codes at low SNR's. The idea of spectral thinning is then formalized in Section IV through the use of random interleaving. Finally, some conclusions are drawn concerning the distance spectrum of Turbo codes and the consequences of this on the design of codes in general.

2 The Free Distance of Turbo Codes

In order to find the free distance of a Turbo code, it is necessary to understand the basic structure of the encoder and the resulting codewords. A typical Turbo encoder consists of the parallel concatenation of two or more, usually identical, rate 1/2 encoders, realized in systematic feedback form, and an interleaver. This encoder structure is called a parallel concatenation because the two encoders operate on the same *set* of input bits, rather than one encoding the output of the other. A block diagram of a Turbo encoder with two constituent convolutional encoders is shown in Figure 2. For the remainder of the paper, only Turbo encoders with two identical constituent convolutional encoders are considered, though the conclusions are easily extended to the general case.

The interleaver is used to permute the input bits such that the two encoders are operating on the same *set* of input bits, but different input *sequences*. Thus, the first encoder receives the input bit x_i and produces the output pair (x_i, y_i^1) while the second encoder receives the input bit x_i' and produces the output pair (x_i', y_i^2). The input bits are grouped into finite length sequences whose length, N, equals the size of the interleaver. Since both the encoders are systematic and operate on the same set of input bits, it is only necessary to transmit the input bits once and the overall code has rate 1/3. In order to increase the overall rate of the code to 1/2, the two parity sequences $\{y^1\}$ and $\{y^2\}$ can be punctured by alternately deleting y^1 and y^2. We will refer to a Turbo code whose constituent encoders have parity check polynomials h_0 and h_1, expressed in either octal or D transform notation, and whose interleaver is of length N, as an (h_0, h_1, N) Turbo code.

For example, consider the Turbo encoder shown in Figure 2, where each constituent encoder is a $(2, 1, 2)$ encoder with parity check polynomials $h_0(D) = 1 + D^2$ and $h_1(D) =$

D. For purposes of illustration, assume a pseudorandom interleaver of size $N = 16$ bits which generates a $(1 + D^2, D, 16)$ Turbo code. The interleaver is realized as a 4x4 matrix which is filled sequentially, row by row, with the input bits x_i. Once the interleaver has been filled, the input sequence to the second encoder, x'_i, is obtained by reading the interleaver in a pseudorandom manner until each bit has been read once and only once. The pseudorandom nature of the interleaver in this example is represented by a permutation $\Pi_{16} = \{6, 14, 4, 7, 11, 8, 3, 5, 9, 13, 0, 2, 12, 1, 10, 15\}$, which implies $x'_0 = x_{15}$, $x'_1 = x_{10}$, and so on.

If the input sequence is $x = \{x_{15} \ldots x_0\} = \{0,1,0,1,0,0,0,0,0,0,0,1,0,0,0,1\}$ and the interleaver is represented by the permutation Π_{16}, then the input sequence to the second encoder is $x' = \Pi_{16}(x) = \{0,1,1,0,0,0,0,0,0,0,1,0,1,0,0,0\}$. The trellis diagrams for both constituent encoders with these inputs are shown in Figure 3. The corresponding unpunctured parity sequences are $y^1 = \{0,0,1,0,0,0,0,0,0,0,0,0,1,0,1,0\}$ and $y^2 = \{1,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0\}$. The resulting codeword has Hamming weight $d = w(x) + w(y^1) + w(y^2) = 4 + 3 + 3 = 10$ without puncturing, where $w(x)$ is the Hamming weight of the sequence x. If the code is punctured beginning with y^1_0, then the resulting codeword has weight $d = 4 + 3 + 2 = 9$. If, on the other hand, the puncturing begins with y^2_0, then the punctured codeword has Hamming weight $4 + 1 + 1 = 6$.

Finding the free distance of a Turbo code is complicated by the fact that Turbo encoders are time-varying due to the interleaver. That is, if $\tilde{x} = Dx$, where D is the delay operator [45], then $\tilde{y}^1 = Dy^1$, but $\tilde{x}' \neq Dx'$ (with high probability) and $\tilde{y}^2 \neq Dy^2$. (Here, we only consider delays of a finite length sequence x for which no ones are lost.) Continuing the example, if $\tilde{x} = Dx$, then $\tilde{x}' = \Pi_{16}(\tilde{x}) = \{0,0,0,0,0,0,0,1,0,1,0,0,0,1,0,1\}$ and $\tilde{y}^2 = \{0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0\}$. Thus, time shifting the input bits results in codewords that differ in both bit position and overall Hamming weight! The variation in the weights of codewords corresponding to time shifted input sequences is magnified by puncturing.

This simple example illustrates several salient points concerning the structure of the codewords. First, because the pseudorandom interleaver permutes the input bits, the two input sequences x and x' are almost always different, though of the same weight, and the two encoders will (with high probability) produce parity sequences of different weights. Second, it is easily seen that a codeword may consist of a number of distinct error events in each encoder, where an error event is a path in the trellis that diverges from the all zero state and then remerges with the all zero state within a finite number of branches. Note that since the constituent encoders are realized in systematic feedback form a nonzero sequence is required to return the encoder to the all zero state. Thus, since at least one nonzero bit is required to start the error event, all error events are associated with information sequences of weight 2 or greater [9]. Finally, with a pseudorandom interleaver it is highly unlikely that both encoders will be returned to the all zero state at the end of the codeword even when the last ν bits of the input sequence x are used to force the first encoder back to the all zero state.

If neither encoder is forced back to the all zero state, i.e., no tail is used, then the sequence consisting of $N-1$ zeroes followed by a one is a valid input sequence x to the first encoder. For some interleavers, this x will be permuted to itself and x' will be the same sequence. In this case, with puncturing, the weight of the codeword and the free distance of the code will be at most two. Note that this codeword is caused by an information sequence of weight one! Thus, forcing the first encoder to return to the all zero state insures that every information sequence has at least weight two and eliminates the possibility of a weight two codeword. For this reason, it is common to assume that the first encoder is forced to return to the all zero state.

The ambiguity of the final state of the second encoder has been shown by simulation to result in negligible performance degradation [13], [31] for large interleavers. For these reasons, it will be assumed for the remainder of the paper that the first encoder is forced to return to the all zero state and that the final state of the second encoder is unknown. Special interleaver structures that result in both encoders returning to the all zero state are discussed in [32], [33], [35], [36] and [37].

2.1 Performance Bounds

In order to make clear the distinction between Turbo codes and convolutional codes, it is useful to consider these codes as block codes. To this end, the input sequences are restricted to length N, where N corresponds to the size of the interleaver in the Turbo encoder. With finite length input sequences of length N, a $(2,1,\nu)$ convolutional code may be viewed as a block code with 2^N codewords of length $2(\nu+N)$.

The bit error rate (BER) performance of a convolutional code with maximum likelihood (ML) decoding on an additive white Gaussian noise (AWGN) channel with an SNR of E_b/N_0 can be upper bounded using a union bound technique by [13]

$$P_b \leq \sum_{i=1}^{2^N} \frac{w_i}{N} Q\left(\sqrt{d_i \frac{2RE_b}{N_0}}\right), \tag{1}$$

where w_i and d_i are the information weight and total Hamming weight, respectively, of the i^{th} codeword. Collecting codewords of the same total Hamming weight and defining the average information weight per codeword as

$$\bar{w}_d = \frac{W_d}{N_d},$$

where W_d is the total information weight of all codewords of weight d and N_d is the total number, or multiplicity, of codewords of weight d, yields

$$P_b \leq \sum_{d=d_{free}}^{2(\nu+N)} \frac{N_d \bar{w}_d}{N} Q\left(\sqrt{d\frac{2RE_b}{N_0}}\right), \tag{2}$$

where d_{free} is the free distance of the code. (In this development, the multiplicity N_d^0 includes codewords due to multiple error events for $d \geq 2d_{free}$).

If a convolutional code has N_d^0 codewords of weight d caused by information sequences x whose first one occurs at time 0, then it also has N_d^0 codewords of weight d caused by the information sequences Dx, N_d^0 codewords of weight d caused by the information sequences D^2x, and so on. Thus, as the length of the information sequences increases, we have

$$\lim_{N\to\infty} \frac{N_d}{N} = N_d^0$$

and

$$\lim_{N\to\infty} \bar{w}_d = \lim_{N\to\infty} \frac{W_d}{N_d} = \frac{W_d^0}{N_d^0} \triangleq \bar{w}_d^0,$$

where W_d^0 is the total information weight of all codewords with weight d which are caused by information sequences whose first one occurs at time 0. Thus, the bound on the BER of a convolutional code with ML decoding becomes

$$P_b \leq \sum_{d=d_{free}}^{2(\nu+N)} N_d^0 \bar{w}_d^0 Q\left(\sqrt{d\frac{2RE_b}{N_0}}\right) = \sum_{d=d_{free}}^{2(\nu+N)} W_d^0 Q\left(\sqrt{d\frac{2RE_b}{N_0}}\right), \tag{3}$$

which is the standard union bound for ML decoding. For this reason, efforts to find good convolutional codes for use with ML decoders have focused on finding codes that maximize the free distance d_{free} and minimize the number of free distance paths N_{free}^0 for a given rate and constraint length.

The performance of a Turbo code with maximum likelihood decoding can also be bounded using the union bound of equation (2). However, in the Turbo encoder the pseudorandom interleaver maps the input sequence x to x' and the input sequence Dx to a sequence x'' that is different from Dx' with high probability. Thus, unlike convolutional codes, the input sequences x and Dx produce different codewords with different Hamming weights. For Turbo codes with pseudorandom interleavers, $N_d\tilde{w}_d$ is much less than N for low weight codewords. This is due to the pseudorandom interleaver which, with high probability, maps low weight parity sequences in the first constituent encoder to high weight parity sequences in the second constituent encoder. Thus, for low weight codewords

$$\frac{\tilde{w}_d N_d}{N} << 1,$$

where

$$\frac{N_d}{N} \tag{4}$$

is called the *effective multiplicity* of codewords of weight d. The effect of the interleaver size on the multiplicity is also reported in [7], [8], and [9].

2.2 Asymptotic Performance

For moderate and high signal-to-noise ratios, it is well known that the free distance term in the union bound on the bit error rate performance dominates the bound [45]. Thus, for Turbo codes the asymptotic performance approaches

$$P_b \approx \frac{N_{free}\tilde{w}_{free}}{N} Q\left(\sqrt{d_{free}\frac{2RE_b}{N_0}}\right), \tag{5}$$

where N_{free} is the error coefficient and $\tilde{w}_{free}$ is the average weight of the information sequences causing free distance codewords. The expression on the right hand side of equation (5) and its associated graph is called the *free distance asymptote*, P_{free}, of a Turbo code.

An algorithm for finding the free distance of Turbo codes is described in [29]. This algorithm was applied to a Turbo code with the same constituent encoders, puncturing pattern, and interleaver size N as in [1] and a *particular* pseudorandom interleaving pattern. The parity check polynomials for this code are $h_0 = D^4 + D^3 + D^2 + D + 1$ and $h_1 = D^4 + 1$, or $h_0 = 37$ and $h_1 = 21$ using octal notation. This (37,21,65536) code was found to have $N_{free} = 3$ paths with weight $d_{free} = 6$. Each of these paths was caused by an input sequence of weight 2 and thus $\tilde{w}_{free} = 2$. Though this result was for a particular pseudorandom interleaver, it is true for most pseudorandom interleavers with $N = 65536$. This is consistent with the conclusions in [6] in which the performance of Turbo codes is averaged over all possible pseudorandom interleavers.

For this particular Turbo code, the free distance asymptote is given by

$$P_{free} = \frac{3 \times 2}{65536} Q\left(\sqrt{6\frac{E_b}{N_0}}\right),$$

where the rate loss due to the addition of a 4-bit tail is ignored and

$$\frac{N_{free}}{N} = \frac{3}{65536}$$

is the effective multiplicity. The free distance asymptote is shown plotted in Figure 4 along with simulation results for this code using the iterative decoding algorithm of [1] with 18 iterations. From Figure 4, it can clearly be seen that the simulation results do in fact approach the free distance asymptote for moderate and high SNR's. Since the slope of the asymptote is essentially determined by the free distance of the code, it can be concluded that the "error floor" observed with Turbo codes is due to the fact that they have a relatively small free distance and consequently a relatively flat free distance asymptote.

Further examination of equation (5) reveals that the manifestation of the "error floor" can be manipulated in two ways. First, increasing the length of the interleaver while preserving the free distance and the error coefficient will lower the asymptote without changing its slope by reducing the effective multiplicity. In this case, the performance curve of Turbo codes does not flatten out until higher SNR's and lower BER's are reached. Conversely, decreasing the interleaver size while maintaining the free distance and error coefficient results in the error floor being raised and the performance curve flattens at lower SNR's and higher BER's. This can be seen in the simulation results shown in Figure 5 for the $(37, 21, N)$ Turbo code with varying N. If the first constituent encoder is not forced to return to the all zero state and the weight 2 codewords mentioned earlier are allowed, then the error floor is raised to the extent that the code performs poorly even for large interleavers. Thus, one cannot completely disregard free distance when constructing Turbo codes.

If the size of the interleaver is fixed, then the "error floor" can be modified by increasing the free distance of the code while preserving the error coefficient. This has the effect of changing the slope of the free distance asymptote. That is, increasing the free distance increases the slope of the asymptote and decreasing the free distance decreases the slope of the asymptote. It has been shown in [8], [25], [29] and [30] that for a fixed interleaver size, choosing the feedback polynomial to be a primitive polynomial results in an increased free distance and thus a steeper asymptote. An argument to support the use of primitive polynomials in Turbo codes is presented in section 4.

2.3 Comparison to the (2,1,14) Code

The role that the free distance and effective multiplicity play in determining the asymptotic performance of a Turbo code is further clarified by examining the asymptotic performance of a convolutional code. The free distance asymptote of a convolutional code is given by the first term in the union bound of equation (3). The maximum free distance $(2,1,14)$ code whose performance is shown in Figure 4 has $d_{free} = 18$, $N^0_{free} = 18$, and $W^0_{free} = 137$ [46]. Thus, the free distance asymptote for this code is

$$P_{free} = 137\, Q\left(\sqrt{18\frac{E_b}{N_0}}\right),$$

which is also shown in Figure 4.

As expected, the free distance asymptote of the $(2,1,14)$ code is much steeper than the free distance asymptote of the Turbo code due to the increased free distance. However, because the effective multiplicity of the free distance codewords, given by (4), of the Turbo code is much smaller than the multiplicity of the $(2,1,14)$ code, the two asymptotes do not cross until an $E_b/N_0 = 3.5$ dB. At this E_b/N_0, the BER of both codes is less than 10^{-7} which is lower than the targeted BER of many practical systems. Thus, even though the $(2,1,14)$ convolutional code is asymptotically better than the $(37,21,65536)$ Turbo code, the Turbo code is better for the error rates at which many systems operate.

2.4 A Turbo Code with a Rectangular Interleaver

To emphasize the importance of using a pseudorandom interleaver with Turbo codes, we now consider a Turbo code with a rectangular interleaver. Turbo codes with rectangular interleavers have also been considered in [2] where the effect of the interleaver on the free distance of the code is discussed. The same constituent encoders and puncturing pattern as in [1] are used in conjunction with a 120×120 rectangular interleaver. This rectangular interleaver is realized as a 120×120 matrix into which the information sequence x is written row by row. The input sequence to the second encoder x' is then obtained by reading the matrix column by column. A 120×120 rectangular interleaver implies an interleaver size of $N = 14400$ and thus this is a $(37,21,14400)$ Turbo code.

Using the algorithm described in [29], this code was found to have a free distance of $d_{free} = 12$ with a multiplicity of $N_{free} = 28,900$. For this code, each of the free distance paths is caused by an information sequence of weight 4, so $\tilde{w}_{free} = 4$. The free distance asymptote for this code is thus given by

$$P_{free} = \frac{28900 \times 4}{14400} Q\left(\sqrt{12\frac{E_b}{N_0}}\right).$$

The free distance asymptote is plotted in Figure 6 along with simulation results using the iterative decoding algorithm of [1] with 18 iterations. This figure clearly shows that the free distance asymptote accurately estimates the performance of the code for moderate and high E_b/N_0's.

This code achieves a bit error rate of 10^{-5} at an E_b/N_0 of 2.7dB and thus performs 2 dB worse than the $(37, 21, 65536)$ Turbo code with a pseudorandom interleaver even though it has a much larger free distance. The relatively poor performance of the $(37, 21, 14400)$ Turbo code with a rectangular interleaver is due to the large multiplicity of d_{free} paths. This results in an effective multiplicity of

$$\frac{N_{free}}{N} = \frac{28900}{14400} \approx 2,$$

which is much larger than the effective multiplicity of the $(37, 21, 65536)$ Turbo code. We now show that the large multiplicity is a direct consequence of the use of the rectangular interleaver and that, furthermore, increasing the size of the interleaver does not result in a significant reduction in the effective multiplicity of the free distance codewords.

The free distance paths in the Turbo code with the rectangular interleaver are due to four basic information sequences of weight 4. These information sequences are depicted in Figure 7 as they would appear in the rectangular interleaver. The "square" sequence in Figure 7a depicts the sequence $x = 1, 0, 0, 0, 0, 1, 0_{594}, 1, 0, 0, 0, 0, 1, 0_{\infty}$, where 0_{594} denotes a sequence of 594 consecutive zeroes and 0_{∞} represents a sequence of zeroes that continues to the end of the information sequence. In this case, the rectangular interleaver maps the sequence x to itself and therefore $x' = x$. The sequence x results in a parity sequence y^1 from the first constituent encoder which, after puncturing, has weight 4. Similarly, the input sequence $x' = x$ results in a parity sequence y^2 from the second constituent encoder which, after puncturing, also has weight 4. The weight of the codeword is then $d_{free} = 4 + 4 + 4 = 12$. By counting the number of distinct positions in which these sequences can appear in the rectangular interleaver, we can find the multiplicity of the free distance codewords. Since the "square" sequence in Figure 7a can appear in $(\sqrt{N} - 5) \times (\sqrt{N} - 5) = 13,225$ distinct positions in the rectangular interleaver, and in each case $x' = x$ and a codeword of weight $d_{free} = 12$ results, this results in 13,325 free distance codewords. Note that for every occurrence of the "square" sequence to result in a codeword of weight 12 the weight of both parity sequences must be invariant to which is punctured first.

The "rectangular" sequences in Figures 7b and 7c also result in weight 12 codewords. For these two sequences, the weight of one of the parity sequences is affected by whether or not it is punctured first and only every other position in which the "rectangular" sequences appear in the interleaver results in a codeword of weight $d_{free} = 12$. Thus, the sequences in Figure 7b and Figure 7c each result in $0.5(\sqrt{N}-10) \times (\sqrt{N}-5) = 6,325$ free distance codewords. For the sequence in Figure 7d, the weight of both parity sequences is affected by which is punctured first and only one out of four positions in which the "rectangular" sequence appears in the interleaver results in a codeword of weight $d_{free} = 12$. Consequently, this sequence results in $0.25(\sqrt{N}-10) \times (\sqrt{N}-10) = 3,025$ free distance codewords. Summing the contributions of each type of sequence results in a total of $N_{free} = 28,900$ codewords of weight $d_{free} = 12$.

It is tempting to try to improve the performance of a Turbo code with a rectangular interleaver by increasing the size of the interleaver. However, all of the information sequences shown in Figure 7 would still occur in a larger rectangular interleaver, so the free distance cannot be increased by increasing N. Also, since the number of free distance codewords is on the order of N, increasing the size of the interleaver results in a corresponding increase in N_{free} such that the effective multiplicity N_{free}/N does not change significantly. Without the benefit of a reduced effective multiplicity, the free distance asymptote, and thus the "error floor", of Turbo codes with rectangular interleavers is not lowered enough for them to manifest the excellent performance of Turbo codes with pseudorandom interleavers for moderate BER's. Attempts to design interleavers for Turbo codes generally introduce structure to the interleaver and thus destroy the very randomness that results in such excellent performance at low SNR's.

3 The Distance Spectrum of Turbo Codes

In the previous section, it was shown that the "error floor" observed in the performance of Turbo codes is due to their relatively low free distance. It is now shown that the outstanding performance of Turbo codes at low SNR's is a manifestation of the sparse distance spectrum that results when a pseudorandom interleaver is used in a parallel concatenation scheme. To illustrate this the distance spectrum of an "average" (37, 21, 65536) Turbo code is found and its relationship to the performance of the code is discussed. The distance spectrum of the "average" Turbo code is then compared to the distance spectrum of the (2, 1, 14) code. An "average" Turbo code is one whose properties have been averaged over all possible pseudorandom interleavers [8]. By doing this, the analysis of Turbo codes and the algorithm for finding the distance spectrum are simplified.

Using the algorithm described in [29], an "average" (37, 21, 65536) Turbo code was found to have the following distance spectrum

d	N_d	W_d
6	3	6
8	22	54
10	41	157
12	150	323

where N_d is the total number of codewords of weight d and $W_d = N_d\tilde{w}_d$ is the total information weight of all codewords of weight d. The distance spectrum information for a particular distance d is referred to as a spectral line. This data can be used in conjunction with the bound of equation (2) to estimate the performance of the code. In addition, by plotting each term of equation (2) the contribution of each spectral line to the overall performance of the code can be estimated.

The performance of a (37, 21, 65536) Turbo code is shown in Figure 8 along with curves showing the contribution of each spectral line for an "average" Turbo code with the same interleaver length. This clearly shows that the contribution to the code's BER by the higher distance spectral lines is less than the contribution of the free distance term for E_b/N_0's greater than 0.5 dB. Thus, the free distance asymptote dominates the performance of the code not only for moderate and high E_b/N_0's, but also for low E_b/N_0's. We characterize distance spectra for which this is true as sparse or spectrally thin.

3.1 Comparison to the (2,1,14) Code

The ramifications of a sparse distance spectrum are made evident by examining the distance spectrum of convolutional codes. The (2, 1, 14) convolutional code introduced in section 2, has the following distance spectrum

d	N_d^0	W_d^0
18	33	187
20	136	1034
22	835	7857
24	4787	53994
26	27941	361762
28	162513	2374453
30	945570	15452996
32	5523544	99659236

as reported in [46]. When comparing the distance spectrum of a convolutional code and a Turbo code, it is important to remember that for a convolutional code $N_d \approx N \times N_d^0$ for the low weight codewords. Figure 9 shows the estimated performance of this code using the bound of equation (3) and the contribution of each spectral line.

In this case, the contribution of the higher distance spectral lines to the overall BER is greater than the contribution of the free distance term for E_b/N_0's less than 2.7 dB, which corresponds to BER's of less than 10^{-6}! The large SNR required for the free distance asymptote to dominate the performance of the $(2, 1, 14)$ code is due to the rapid increase in the path multiplicity for increasing d. We characterize distance spectra for which this is true as spectrally dense. The dense distance spectrum of convolutional codes also accounts for the discrepancy between the real coding gain at a particular SNR and the asymptotic coding gain calculated using just the free distance [45].

Thus, it can be concluded that the outstanding performance of Turbo codes at low signal-to-noise ratios is a result of the dominance of the free distance asymptote, which in turn is a consequence of the sparse distance spectrum of Turbo codes, as opposed to spectrally dense convolutional codes. Finally, the sparse distance spectrum of Turbo codes is due to the structure of the codewords in a parallel concatenation and the use of pseudorandom interleaving.

4 Spectral Thinning

In this section, the observations made concerning the distance spectrum and spectral thinning of Turbo codes is formalized from the point of view of random interleaving. Random interleaving was introduced in [6]–[10] to develop transfer function bounds on the average performance of Turbo codes and to explore issues of code design. Here, random interleaving is used to explore the effect of the interleaver on the distance spectrum of the code. In order to simplify the notation and discussion, only nonpunctured Turbo codes are considered explicitly. The extension to punctured codes is straightforward and may be found in [29].

The fundamental idea of random interleaving is to consider the performance of a Turbo code averaged over all possible pseudorandom interleavers of a given length. For a given N, there are $N!$ possible pseudorandom interleavers and, assuming a uniform distribution, each occurs with probability $\frac{1}{N!}$. Let a particular interleaver map an information sequence x of weight w to an information sequence x', also of weight w. Then there are a total of $w!(N-w)!$ interleavers in the ensemble of $N!$ interleavers that perform this same mapping. Thus, the probability that such a mapping and, hence, that the codeword that results from the input sequences x and x' occurs is

$$\frac{w!(N-w)!}{N!} = \frac{1}{\begin{pmatrix} N \\ w \end{pmatrix}}.$$

Following [8], define the input redundancy weight enumerating function (IRWEF) of a systematic code as

$$A(W,Z) = \sum_{w} \sum_{z} A_{w,z} W^{w} Z^{z}, \tag{6}$$

where $A_{w,z}$ is the number of codewords of weight $d = w + z$ generated by input sequences of weight w and parity sequences of weight z. The goal is now to develop a relationship between the codewords in the constituent encoders and $A_{w,z}$ for the Turbo code and to see how that relationship changes with the size of the interleaver.

Recall from section 2 that a Turbo codeword is essentially the combination of a codeword from the first constituent encoder plus a codeword from the second constituent encoder. A codeword of weight $d_1 = w + z_1$ from the first constituent encoder caused by an information sequence x of weight w is composed of n_1 error events of total length l_1. To avoid difficulties in counting codewords when the second constituent encoder is left unterminated, we consider different orderings of the same error events as distinct cases. The ordered set of n_1 error events in the first encoder is denoted by S_1. The information sequence x that results in the set S_1 is mapped by a particular interleaver to the information sequence x', also of weight w, which is then encoded by the second constituent encoder. This results in a codeword of weight $d_2 = w + z_2$, with the ordered set S_2 consisting of n_2 error events of total length l_2.

For example, Figure 3 depicts a codeword of weight $d_1 = 4 + 3 = 7$ in the first constituent encoder caused by an information sequence of weight $w = 4$ and composed of $n_1 = 2$ error events of total length $l_1 = 5 + 3 = 8$. For the interleaver described in section

2, x is mapped to an x' that results in a codeword of weight $d_2 = 4+3 = 7$ in the second constituent encoder consisting of $n_2 = 2$ error events of total length $l_2 = 3+3 = 6$. Thus, this S_1 and S_2 result in a codeword of weight 10 and a single contribution to $A_{4,6}$ of the Turbo code for this particular interleaver. Averaged over the ensemble of interleavers of length N, a set S_1 with information weight w and parity weight z_1 and a set S_2 with information weight w and parity weight z_2 will contribute a fraction

$$\frac{1}{\binom{N}{w}}, \tag{7}$$

to the enumerating function coefficients A_{w,z_1+z_2} of an "average" Turbo code.

Because the sequence of zeroes connecting any two distinct error events has no effect on the weight of the information sequence or the parity sequence, there are

$$\binom{N - l_1 + n_1}{n_1} \tag{8}$$

ways that the ordered set, S_1, of n_1 error events can be arranged such that their contribution to A_{w,z_1+z_2} is not changed. This is simply the number of ways in which n_1 distinct error events can be arranged in a sequence of length N while maintaining the order in which they appear. Similarly, if the codeword in the second constituent encoder ends in the all zero state, then the ordered set S_2 will make

$$\binom{N - l_2 + n_2}{n_2} \tag{9}$$

contributions to A_{w,z_1+z_2}. However, because the second encoder is not guaranteed to return to the all zero state, it is possible that the last of the n_2 error events is not actually an error event, but instead ends in a nonzero state. In this case, the last error event cannot be moved and the set S_2 makes

$$\binom{N - l_2 + (n_2 - 1)}{(n_2 - 1)} \tag{10}$$

contributions to A_{w,z_1+z_2}.

The contribution to the distance spectrum of a Turbo code due to any pair of ordered sets S_1 and S_2, averaged over the ensemble of pseudorandom interleavers, can now be computed using equations (7), (8), (9), and (10). If the codeword in the second constituent encoder happens to end in the all zero state, then the contribution of some S_1 and S_2 to A_{w,z_1+z_2} is given by

$$\frac{\binom{N-l_1+n_1}{n_1}\binom{N-l_2+n_2}{n_2}}{\binom{N}{w}}. \tag{11}$$

If the codeword in the second constituent encoder does not end in the all zero state, then the contribution to A_{w,z_1+z_2} is given by

$$\frac{\binom{N-l_1+n_1}{n_1}\binom{N-l_2+(n_2-1)}{(n_2-1)}}{\binom{N}{w}}. \tag{12}$$

Equations (11) and (12) can now be used to explore the effect of changing the interleaver size on the distance spectrum of an "average" Turbo code.

Since we are primarily concerned with low weight codewords in the distance spectrum, we assume that $N \gg n_1, n_2, l_1,$ and l_2. If this is not true, then S_1 and S_2 either contain a very large number of short error events or a few very long error events. In both cases, it is very unlikely that the result is a codeword of low weight. With this assumption, equation (11) can be approximated by

$$\frac{w!}{n_1!n_2!}N^{n_1+n_2-w}, \tag{13}$$

where, without loss of generality, it is assumed that $n_1 \geq n_2$. Since each error event is caused by an information sequence of weight at least two, $w \geq 2n_1$. The behavior of equation (13) for increasing N can be broken down into three cases:

1. $n_1 > n_2$
   The exponent of N is strictly negative and the contribution to A_{w,z_1+z_2} decreases as N increases.

2. $n_1 = n_2$ and $w > 2n_1$
   The exponent of N is strictly negative and the contribution to A_{w,z_1+z_2} decreases as N increases.

3. $n_1 = n_2$ and $w = 2n_1$
   The exponent of N is exactly zero and the contribution to A_{w,z_1+z_2} converges to a finite value as N increases.

For $N \gg n_1, n_2, l_1,$ and l_2, equation (12) can be approximated by

$$\frac{w!}{n_1!(n_2-1)!} N^{n_1+n_2-w-1}, \tag{14}$$

where $w \geq 2n_1$. However, since the tail in the second encoder may be caused by an information sequence of weight 1, we also have $w \geq 2n_2 - 1$. The behavior of equation (14) for increasing N can be broken down into three cases:

1. $n_1 > n_2$
   Since $w \geq 2n_1$, the exponent of N is strictly negative and the contribution to A_{w,z_1+z_2} decreases as N increases.

2. $n_1 = n_2$
   Again, since $w \geq 2n_1$, the exponent of N is strictly negative and the contribution to A_{w,z_1+z_2} decreases as N increases.

3. $n_1 < n_2$
   Since $w \geq 2n_2 - 1$, the exponent of N is strictly negative and the contribution to A_{w,z_1+z_2} decreases as N increases.

The following lemma can now be stated. A similar result was proven in [10].

Lemma 1 *Given a Turbo code based on two systematic feedback encoders and a pseudorandom interleaver of length N in which the first encoder is assumed to be forced back to the all zero state, the contribution of two ordered sets of error events S_1 and S_2 with the same information weight to the distance spectrum of the Turbo code, averaged over all pseudorandom interleavers of length N, converges to a nonzero constant as $N \to \infty$, if and only if,*

1. S_2 leaves the second encoder in the all zero state.

2. S_1 and S_2 contain the same number of error events.

3. Each error event in S_1 and S_2 is caused by a weight two information sequence.

In all other cases, the contribution goes to zero as $N \to \infty$.

For each term $A_{w,z}$ in the input redundancy weight enumerating function of a Turbo code there is a finite number of pairs of sets S_1 and S_2 that contribute to it. If either set contains a long error event, then it is possible that that pair will be excluded for small interleavers. As the interleaver size increases, eventually all pairs of sets will be allowed and any further increase in N will not result in additional pairs of sets contributing to $A_{w,z}$. Also, as $N \to \infty$, $A_{w,z}$ will be determined by pairs of S_1 and S_2 that satisfy the three conditions of Lemma 1 and thus each $A_{w,z}$ will converge to a finite value. Since each spectral line is a finite sum of $A_{w,z}$ terms, each spectral line converges to a finite value as the interleaver size increases.

The convergence of each spectral line to a finite value as the size of the interleaver increases results in spectral thinning. That is, for small interleavers there may be pairs of sets S_1 and S_2 that do not satisfy the conditions of Lemma 1, but which contribute to the multiplicity of a low weight spectral line. As the size of the interleaver increases the number of these aberrational sets decreases until the spectral line reaches it final value as determined by Lemma 1. This process of spectral thinning is represented graphically in Figure 10 which depicts the thinning of low weight codewords as the size of the interleaver increases for hypothetical distance spectra. It is this thinning of the distance spectrum that enables the free distance asymptote of a Turbo code to dominate the performance for low SNR and thus to achieve near capacity performance.

4.1 Primitive Polynomials and Free Distance

We now consider the ramifications of Lemma 1 with respect to the free distance codewords of a Turbo code. That is, what does Lemma 1 imply about the information sequences that generate the free distance codewords in an "average" Turbo code?

For an "average" Turbo code, Lemma 1 states that as the size of the interleaver increases each spectral line is the result of contributions only from pairs of ordered sets of error events in which each error event is caused by a weight two information sequence. It is reasonable to expect that the free distance spectral line will be among the first spectral lines to converge to its final value. Thus, for reasonably large interleavers the free distance will be determined by the sets S_1 and S_2 satisfying the conditions of Lemma 1. Let s_1 and s_2 be any pair of error events caused by a weight two information sequence that results in a minimum weight parity sequence in the first and second constituent encoders, respectively. Note that there may be more than one such pair of minimum weight error events for the constituent encoders.

A free distance codeword in an "average" Turbo code must be the result of sets S_1 and S_2 that consist of only those minimum weight error events s_1 and s_2, respectively.

Furthermore, since each additional error event in either S_1 and S_2 adds weight to the codeword, S_1 and S_2 must each contain only one minimum weight error event. (If each error event does not add weight, then a weight two information sequence exists that generates a zero weight parity sequence and the free distance of the code would be 2.) Thus, the free distance codewords of an "average" Turbo code are caused by weight two information sequences, provided the interleaver is large enough.

Therefore we have the following Lemma.

Lemma 2 *For an "average" Turbo code, as the size of the interleaver N approaches ∞:*

1. *The free distance codewords are caused by information sequences of weight 2.*
2. *The free distance of an "average" Turbo code is maximized by choosing constituent encoders that have the largest output weight for weight two information sequences.*

Based on this lemma, we now present an intuitive argument for why choosing the feedback polynomial $h_0(D)$ in a $(2, 1, \nu)$ systematic feedback encoder to be a primitive polynomial maximizes the output weight for weight two information sequences. It follows that the free distance of an "average" Turbo code is maximized by using a primitive polynomial as the feedback polynomial in the constituent encoders. This result was independently derived in [10] using the transfer function of an "average" Turbo code.

The generator matrix of a $(2, 1, \nu)$ systematic feedback encoder is given by

$$G_{fb}(D) = \begin{bmatrix} 1 & \dfrac{h_1(D)}{h_0(D)} \end{bmatrix},$$

where $h_1(D)$ and $h_0(D)$ are referred to as the feedforward and feedback polynomials, respectively, and $h_0(D)$ is of degree ν. Since only information sequences of weight 2 are being considered, the systematic output contributes weight 2 to the overall codeword weight for all the encoders being considered. Therefore, only the weight contributed by the parity sequence, that is

$$y(D) = x(D)\frac{h_1(D)}{h_0(D)},$$

needs to be maximized. Furthermore, since we are concerned only with the choice of $h_0(D)$, $h_1(D)$ is assumed to be a polynomial such that $h_0(D)$ and $h_1(D)$ are relatively prime. (There is empirical evidence that the choice of both polynomials can affect the performance of the code [29], [31], but we will not address that issue in this paper.)

Let $1+D^K$, for some finite K, be the shortest input sequence of weight 2 that generates a finite length codeword. The resultant parity sequence is

$$\begin{aligned} y(D) &= (1+D^K)\frac{h_1(D)}{h_0(D)} \\ &= \frac{h_1(D)}{h_0(D)} + D^K\frac{h_1(D)}{h_0(D)}. \end{aligned}$$

Since $y(D)$ is of finite length, $h_1(D)/h_0(D)$ must be periodic with period K. Increasing the period K increases the length of the shortest weight 2 input sequence that generates a finite length codeword and therefore increases the length of that codeword. Intuitively, one would expect that increasing its length would result in the codeword gaining weight. That is, on average, half of the added bits would be ones.

A strictly proper rational function of two polynomials, like $h_1(D)/h_0(D)$, is periodic with period $K \leq 2^\nu - 1$. The period is maximized, that is, $K = 2^\nu - 1$, when $h_0(D)$ is a primitive polynomial. Since the free distance of an "average" Turbo code is determined by information sequences of weight 2, for sufficiently large interleavers the free distance will be maximized by maximizing K. Therefore, choosing $h_0(D)$ to be a primitive polynomial will result in a larger free distance for an "average" Turbo code.

To test this, we compare a $(37, 21, 400)$ Turbo code which has $K = 5$ and free distance $d_{free} = 6$ to a $(23, 35, 400)$ Turbo code. Both codes are punctured as in [1]. The feedback polynomial $h_0 = 23$ in the second Turbo code is a primitive polynomial of degree $\nu = 4$ and thus $1/h_0$ has a period of $K = 2^\nu - 1 = 15$. The free distance of this Turbo code was found to be $d_{free} = 10$ [13], [29]. Figure 11 shows simulation results for these two codes using the iterative decoding algorithm of [1] with 18 iterations. As expected, the second Turbo code performs better at moderate and high SNR's because its free distance asymptote is steeper due to the increased free distance.

5 Conclusions

The excellent performance of Turbo codes may be explained in terms of the distance spectrum of the code. The 'error floor' observed in simulations of Turbo codes is a manifestation of the free distance asymptote. Since Turbo codes have relatively low free distances the free distance asymptote has a shallow slope, and thus the performance curves flatten out at moderate to high SNR's. The 'error floor' may be lowered by increasing the size of the interleaver for a fixed free distance, that is, by reducing the effective multiplicity of the code. Alternatively, for fixed interleaver lengths, the performance may be improved for moderate and high SNR's by increasing the free distance. Choosing primitive polynomials as the feedback polynomials in the constituent encoders usually results in an increased free distance.

The exceptional performance of Turbo codes at low SNR's is due to the sparse distance spectrum and the resultant ability of the code to follow the free distance asymptote at moderate to low SNR's. The use of systematic feedback encoders and pseudorandom interleavers results in spectral thinning, in which information sequences which generate low weight parity sequences from the first constituent encoder are interleaved with high probability to information sequences that generate high weight parity sequences in the second constituent encoder. Spectral thinning is enhanced by increasing interleaver lengths. For very large interleavers, spectral thinning results in a sparse distance spectrum in which the first several spectral lines are determined solely by input sequences of weight two. Thus, spectral thinning results in few low weight codewords and a large number of codewords of "average" weight. This is very similar to the type of distance spectrum achieved by "random-like" codes [4].

In a more philosophical light, Turbo codes remind us that information theoretical arguments imply that long block lengths, but not necessarily large free distances, are required to achieve capacity at moderate BER's. Thus, like convolutional codes, Turbo codes are a class of codes that achieve long block lengths, but without the corresponding increased density of the distance spectrum common to convolutional codes, and for which a practical, albeit nontrivial, decoding algorithm exists. In addition, Turbo codes are time-varying due to the pseudorandom interleaver, and the time-varying structure is essential in achieving the distance spectrum that results in near capacity performance at moderate BER's. This suggests that some effort should be made to find other classes of time-varying codes, and decoding algorithms, that have good distance spectra, rather than just large free distances. Finally, since, in fact, long block lengths are required to achieve near capacity performance at moderate BER's, only modest coding gains will be achievable in systems that use relatively short block lengths.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes", Proc. 1993 IEEE Int. Conf. on Comm., Geneva, Switzerland, pp. 1064–1070, 1993.

[2] C. Berrou and A. Glavieux, "Turbo-codes: General Principles and Applications", Proceedings of the 6th Tirrenia International Workshop on Digital Communications, Tirrenia, Italy, pp. 215-226, September 1993.

[3] G. Battail, C. Berrou and A. Glavieux, "Pseudo-Random Recursive Convolutional Coding for Near-Capacity Performance", Proceedings of the Comm. Theory Mini-Conference, Globecom '93, Houston, Texas, pp. 23–27, December 1993.

[4] G. Battail, "On random-like codes", preprint.

[5] G. Battail, "Pseudo-random Turbo-codes", *Proc. Int. Symp. on Signals, Systems and Electronics*, San Francisco, U.S.A., pp. 419–422, October 25–27, 1995.

[6] S. Benedetto and G. Montorsi, "Average performance of parallel concatenated block codes", *Electronics Letters*, Vol. **31**, No. 3, pp. 156–158, February 2, 1995.

[7] S. Benedetto and G. Montorsi, "Performance evaluation of TURBO-codes", *Electronics Letters*, Vol. **31**, No. 3, pg. 163-165, February 2, 1995.

[8] S. Benedetto and G. Montorsi, *Unveiling TURBO-codes: some results on parallel concatenated coding schemes*, Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 32, September 1995.

[9] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes", *submitted to IEEE Trans. Inform. Theory*, Jan. 1995.

[10] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes", *to appear in the IEEE Transactions on Communications.*

[11] J. D. Andersen, *The TURBO Coding Scheme*, Report IT-146, Technical University of Denmark, June 1994.

[12] J. D. Andersen, "TURBO Coding for Deep Space Applications", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 36, September 1995.

[13] P. Robertson, "Illuminating the Structure of Parallel Concatenated Recursive Systematic (TURBO) Codes", Proc. GLOBECOM '94, Vol. 3, pp. 1298–1303, San Francisco, California, November 1994.

[14] J. Haganauer and L. Papke, "Decoding "Turbo"-codes with the Soft Output Viterbi Algorithm" Proceedings of the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, pg. 164, June 1994.

[15] J. Hagenauer, P. Robertson, and L. Papke, "Iterative (TURBO) Decoding of Systematic Convolutional Codes with MAP and SOVA Algorithms", Proceedings of the ITG Conference, Frankfurt, Germany, October 1994.

[16] Y. V. Svirid, "Weight distributions of turbo codes", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 38, September 1995.

[17] Y. V. Svirid, "Weight distributions and bounds for turbo codes", *European Transactions on Telecommunications*, Vol. 6, No. 5, pp. 543–556, September-October 1995.

[18] Y. V. Svirid and S. Riedel, "Threshold decoding of turbo codes", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 39, September 1995.

[19] D. Divsalar and F. Pollara, "Low-rate turbo codes for deep-space communications", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 35, September 1995.

[20] D. Divsalar and F. Pollara, "Turbo codes for PCS applications", Proceedings of the 1995 IEEE International Conference on Communications, Seattle, Washington, June, 1995.

[21] D. Divsalar and F. Pollara, "Turbo codes for deep-space communications", *JPL TDA Progress Report 42-120*, February 15, 1995.

[22] D. Divsalar and F. Pollara, "Multiple turbo codes for deep-space communications", *JPL TDA Progress Report 42-121*, May 15, 1995.

[23] D. Divsalar, S. Dolinar, F. Pollara and R. J. McEliece, "Transfer function bounds on the performance of turbo codes", *JPL TDA Progress Report 42-122*, pp. 44–55, August 15, 1995.

[24] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations", *JPL TDA Progress Report 42-122*, pp. 56–65, August 15, 1995.

[25] D. Divsalar and F. Pollara, "On the design of turbo codes", *JPL TDA Progress Report 42-123*, November 15, 1995.

[26] J.-F. Cheng and R. J. McEliece, "Unit-memory Hamming turbo codes", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 33, September 1995.

[27] R. J. McEliece, E. R. Rodemich and J.-F. Cheng, "The turbo decision algorithm", Proceedings of the 33rd Annual Allerton Conference on Communication, Control and Computing, Monticello, Illinois, pg. 366, October 1995.

[28] D. Divsalar and R. J. McEliece, "The effective free distance of turbo codes", *submitted to Electronics Letters.*

[29] J. Seghers, *On the Free Distance of TURBO Codes and Related Product Codes*, Final Report, Diploma Project SS 1995, Number 6613, Swiss Federal Institute of Technology, Zurich, Switzerland, August 1995.

[30] J. Seghers, L. C. Perez and D. J. Costello, Jr., "On selecting code generators for turbo codes", Proceedings of the 33rd Annual Allerton Conference on Communication, Control and Computing, Monticello, Illinois, pg. 357–361, October 1995.

[31] D. Arnold and G. Meyerhans, *The Realization of the the Turbo-Coding System*, Semester Project Report, Swiss Federal Institute of Technology, Zurich, Switzerland, July 1995.

[32] A. S. Barbulescu and S. S. Pietrobon, "Interleaver design for turbo codes", *Electronics Letters*, Vol. **30**, No. 25, pg. 2107, December 8, 1994.

[33] A. S. Barbulescu and S. S. Pietrobon, "Terminating the trellis of turbo-codes in the same state", *Electronics Letters*, Vol. **31**, No. 1, pp. 22–23, January 5, 1995.

[34] A. S. Barbulescu and S. S. Pietrobon, "Rate compatible turbo codes", *Electronics Letters*, Vol. **31**, No. 7, pp. 535–536, March 30, 1995.

[35] A. S. Barbulescu and S. S. Pietrobon, "Interleaver design for three dimensional turbo codes", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 37, September 1995.

[36] M. C. Reed and S. S. Pietrobon, "Turbo-code termination for short frames", *submitted to Electronics Letters.*

[37] O. Joerssen and H. Meyr, "Terminating the trellis of turbo-codes", *Electronics Letters*, Vol. **30**, No. 16, pp. 1285–1286, August 4, 1994.

[38] P. Jung and M. Naßhan, "Performance evaluation of turbo codes for short frame transmission systems", *Electronics Letters*, Vol. **30**, No. 2, pp. 111–113, January 20, 1994.

[39] P. Jung and M. Naßhan, "Dependence of the error performance of turbo-codes on the interleaver structure in short frame transmission systems", *Electronics Letters*, Vol. **30**, No. 4, pp. 285–288, February 17, 1994.

[40] P. Jung, "Novel low complexity decoder for turbo-codes", *Electronics Letters*, Vol. **31**, No. 2, pp. 86–87, January 19, 1995.

[41] R. Podemski, W. Holubowicz, C. Berrou, and A. Glavieux, "Distance spectrum of the turbo-codes", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 34, September 1995.

[42] R. Podemski, W. Holubowicz, C. Berrou, and G. Battail, "Hamming distance spectra of turbo-codes," *Annales Télécommunic.*, Vol. 50, No. 9–10, pp. 790–797, September-October 1995.

[43] G. Caire, G. Taricco, and E. Biglieri, "On the convergence of the iterated decoding algorithm", Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, British Columbia, Canada, pg. 472, September 1995.

[44] N. Wiberg, H.-A. Loeliger, and R. Kötter, "Codes and iterative decoding on general graphs", *European Transactions on Telecommunications*, Vol. 6, No. 5, pp. 513–525, September-October 1995.

[45] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, New Jersey, 1983.

[46] M. Cedervall and R. Johannesson, "A fast algorithm for computing distance spectrum of convolutional codes," *IEEE Trans. Inform. Theory*, **IT-35**, pp. 1146-1159, November 1989.

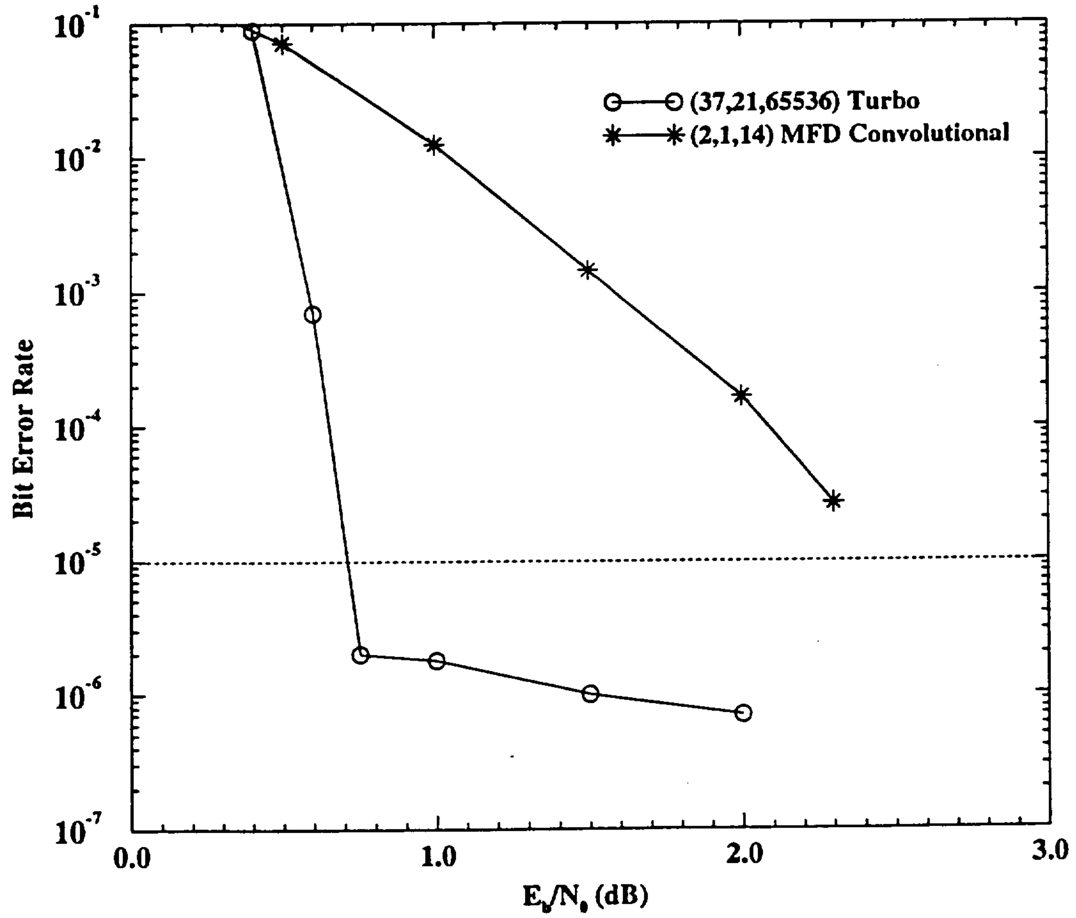


Figure 1: Simulation results for a (37, 21, 65536) Turbo code and a (2, 1, 14) MFD convolutional code.

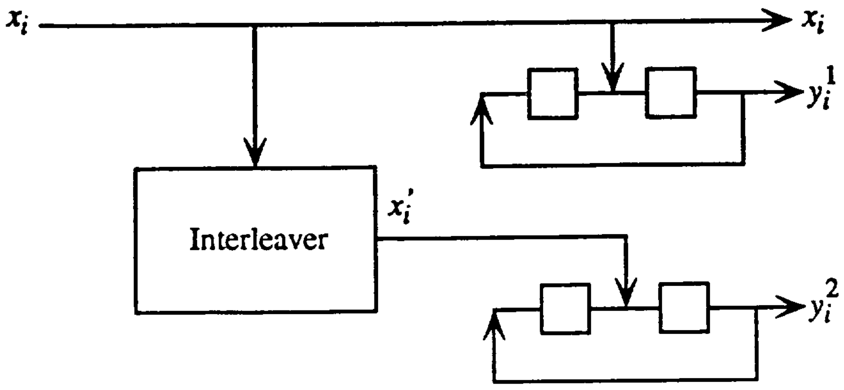


Figure 2: Block diagram of a Turbo encoder with two identical constituent encoders $(h_0(D) = 1 + D^2, h_1(D) = D)$ and without puncturing.

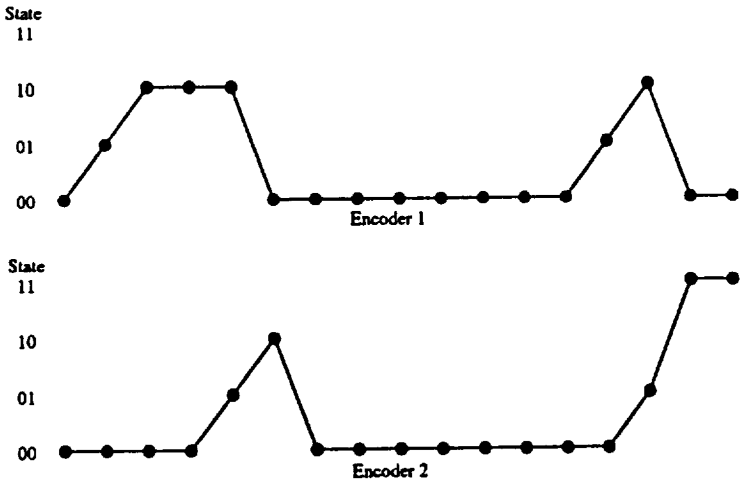


Figure 3: Trellis diagrams for a codeword in the $(1 + D^2, D, 16)$ Turbo code.

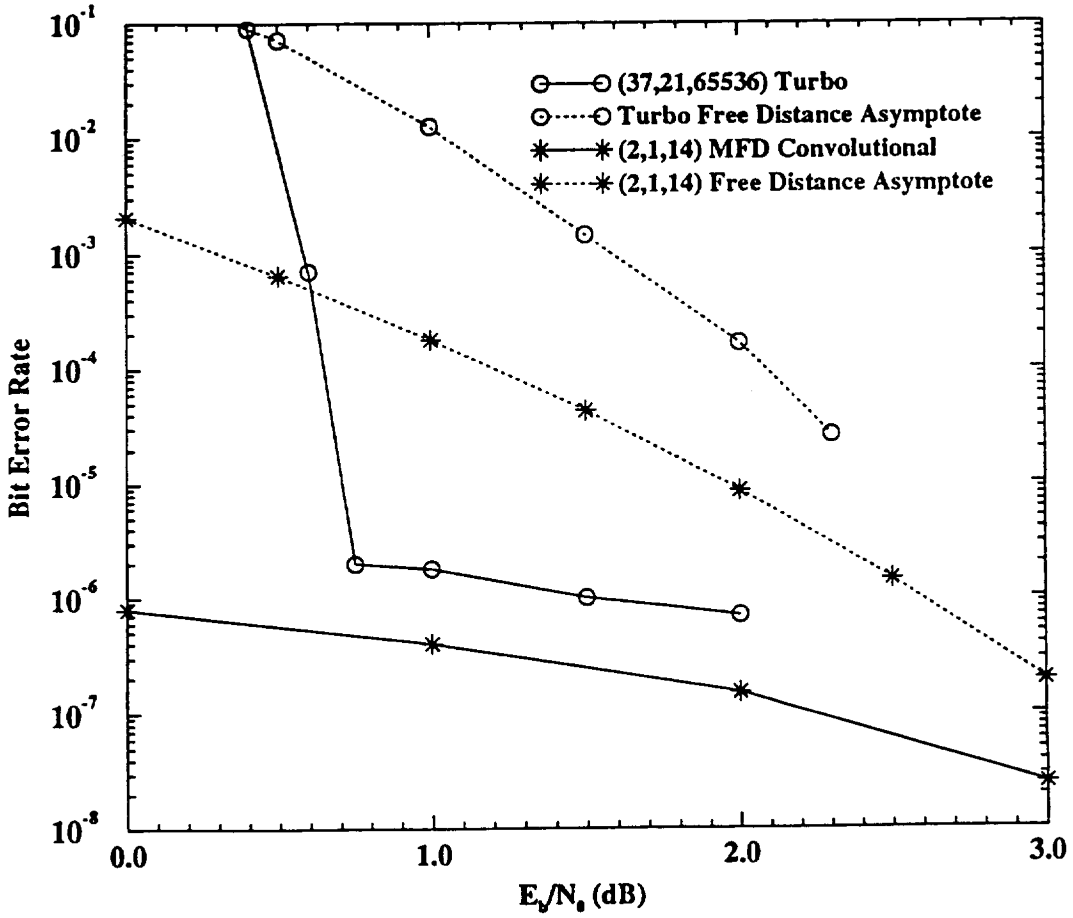


Figure 4: Simulation results for a (37, 21, 65536) Turbo code and a (2, 1, 14) MFD convolutional code and the free distance asymptotes.

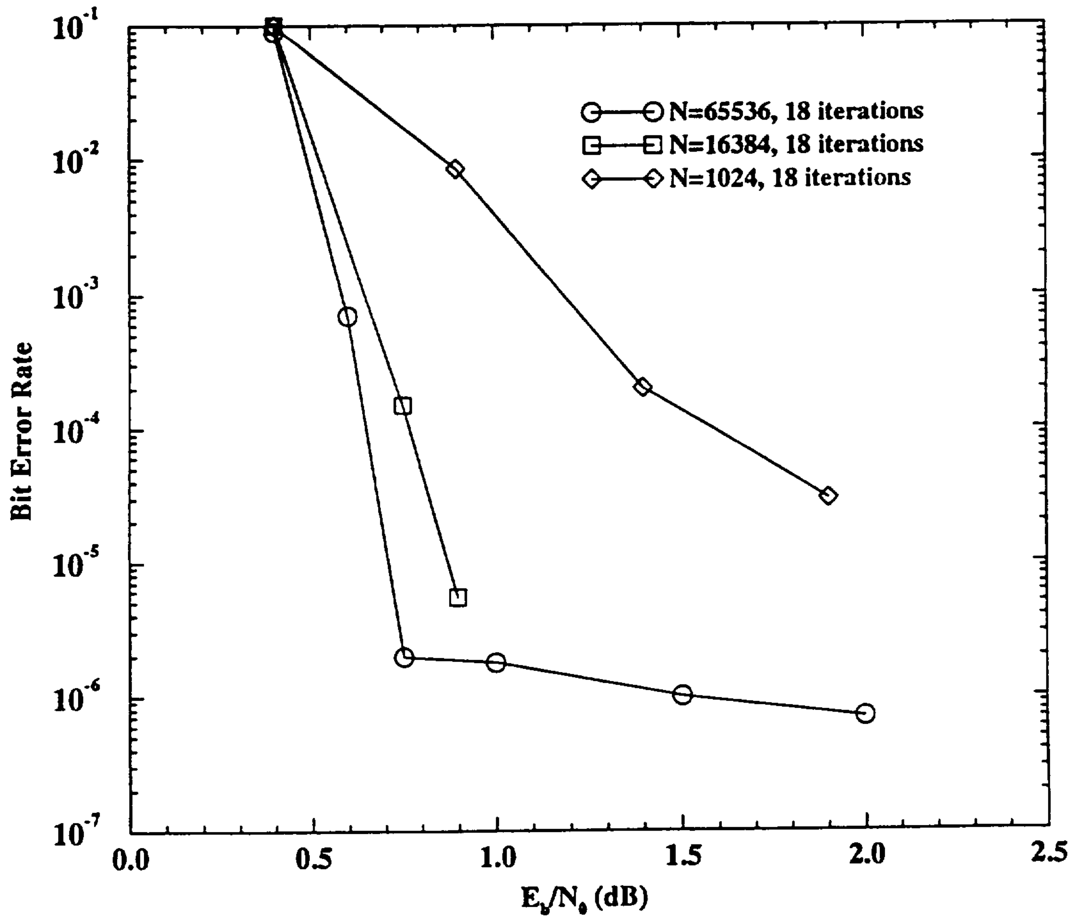


Figure 5: Simulation results for a $(37, 21, N)$ Turbo code with varying interleaver size N.

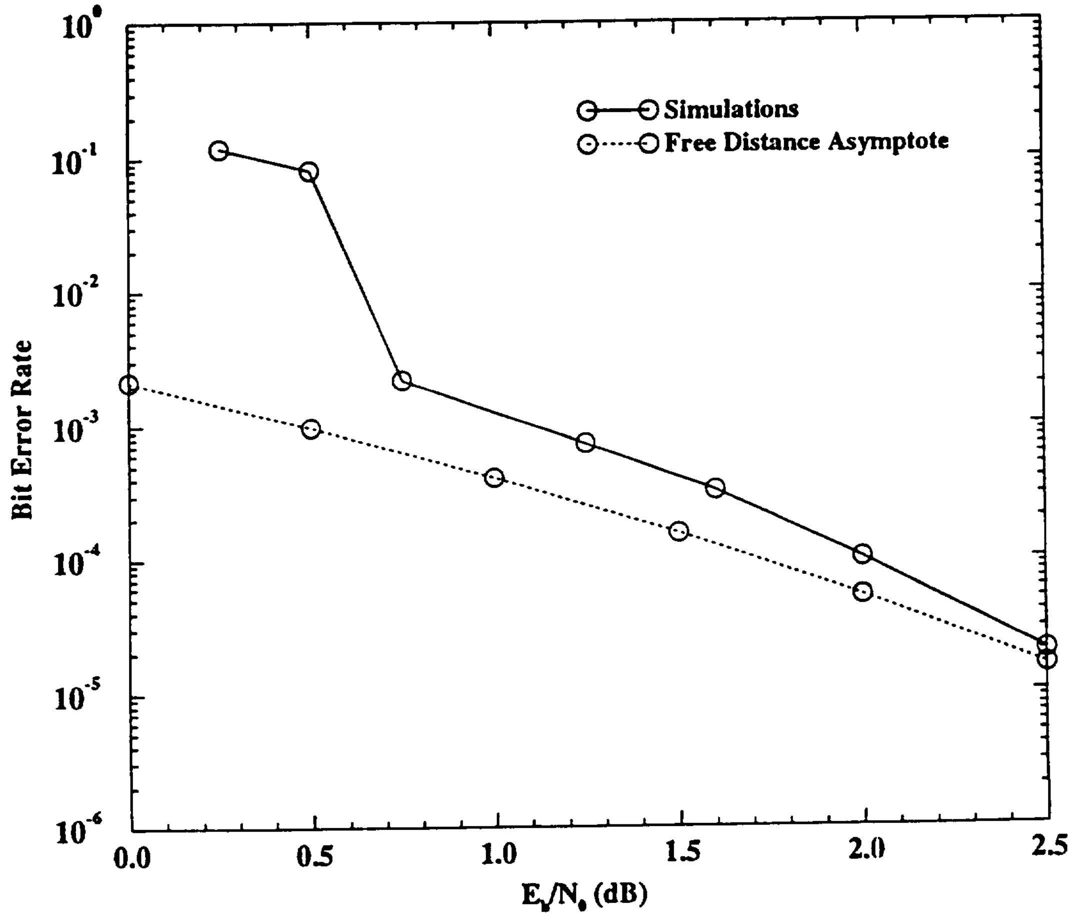


Figure 6: Simulation results and the free distance asymptote for the (37, 21, 14400) Turbo code with a 120 × 120 rectangular interleaver.

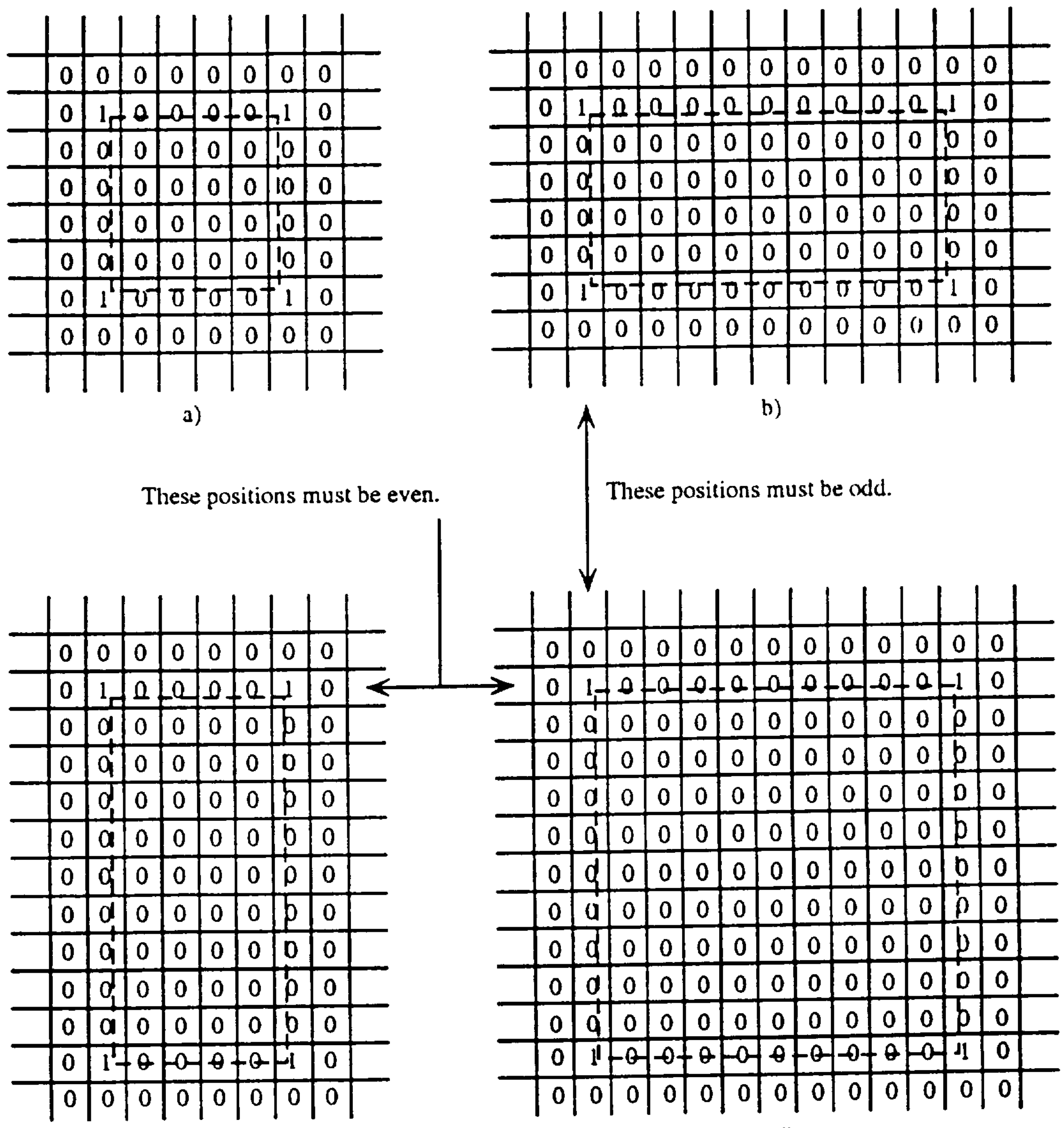


Figure 7: Portions of the interleaver and information sequences resulting in free distance codewords for a (37, 21, 14400) Turbo code with a 120×120 rectangular interleaver.

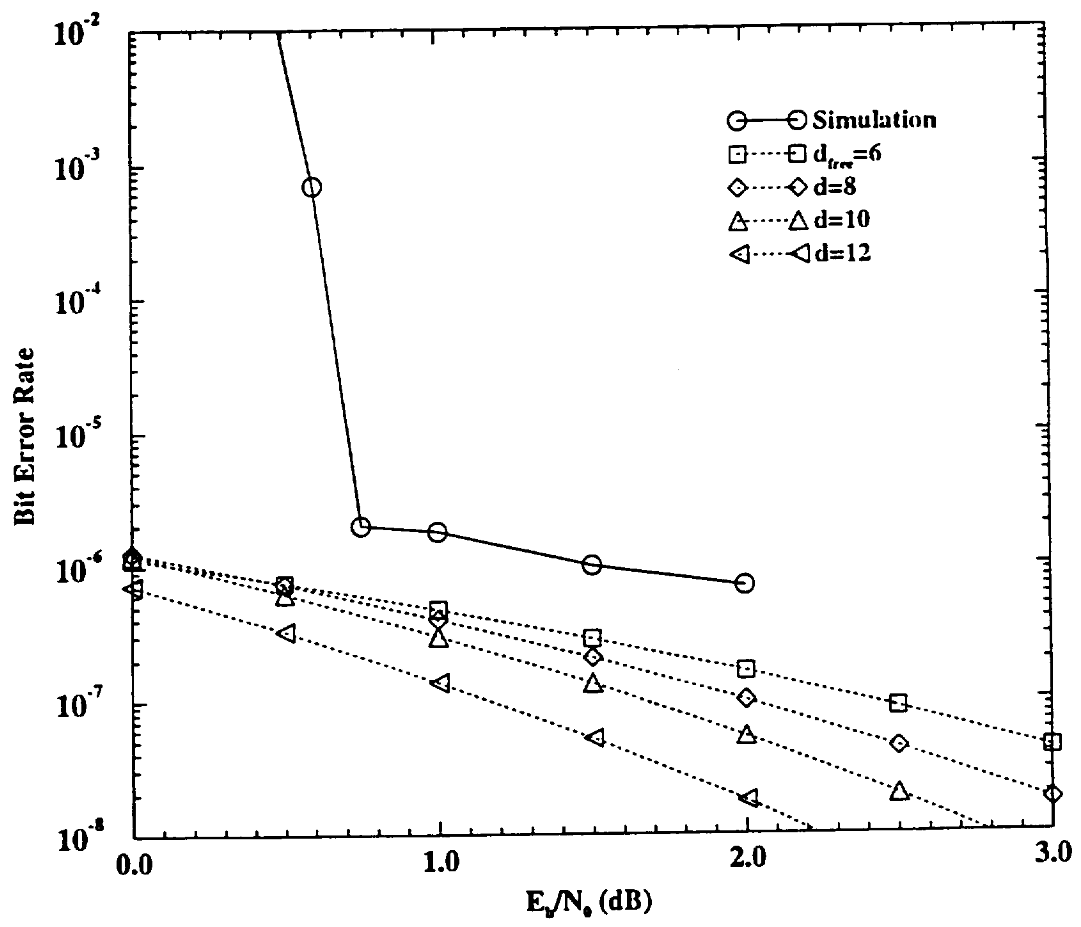


Figure 8: Performance of the (37, 21, 65536) Turbo code decomposed by spectral line.

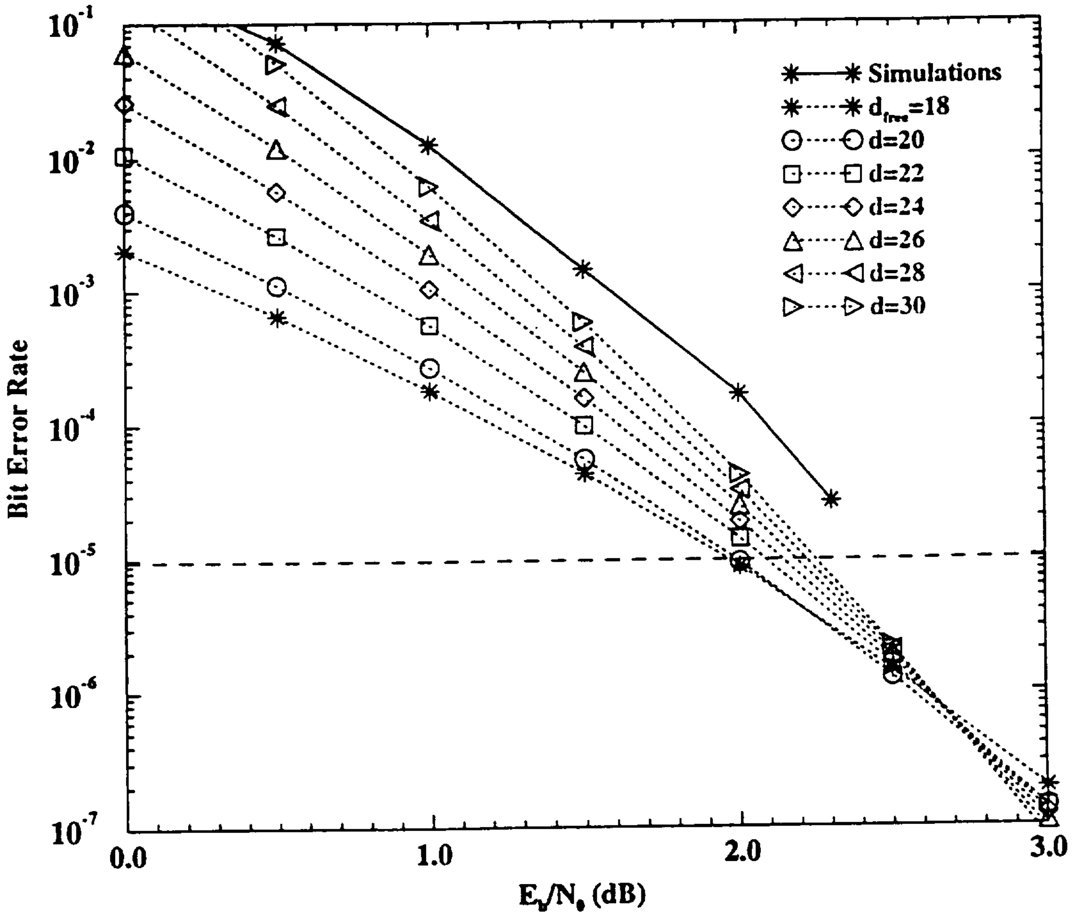


Figure 9: Performance of the (2, 1, 14) MFD convolutional code decomposed by spectral line.

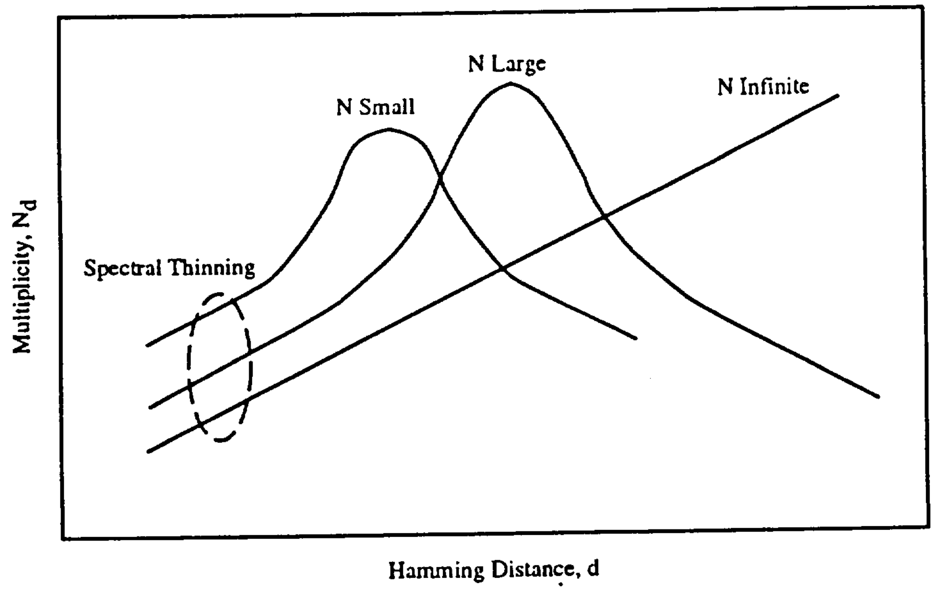


Figure 10: Graphical representation of spectral thinning for increasing interleaver size.

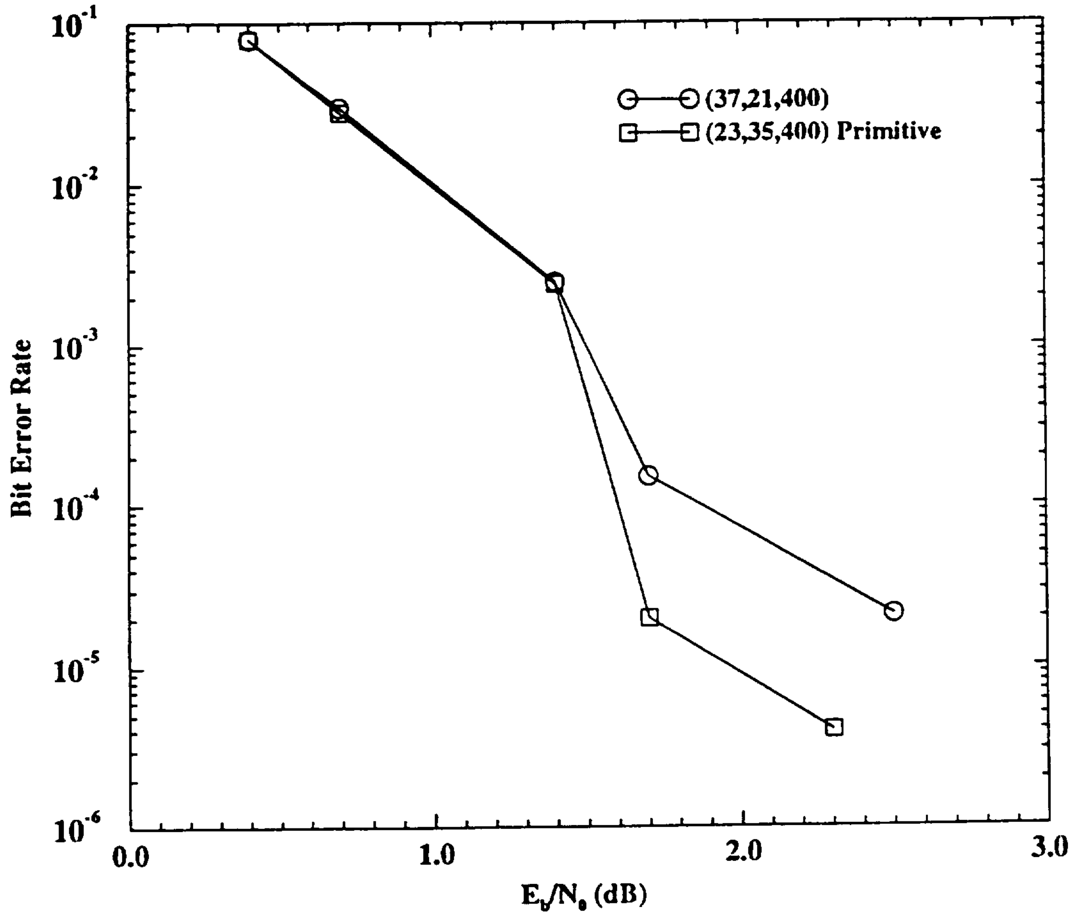


Figure 11: Simulation results for a (37, 21, 400) Turbo code and a (23, 34, 400) primitive Turbo code.